

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

Msrch\_p protein - protein database search, using Smith-Waterman algorithm

Run on: Tue Jan 11 15:44:58 2000; Maspar time 5.30 Seconds  
205,769 Million cell updates/sec

Search output not generated.

Title: >US-09-290-049-2

Description: (1-20) From US09290049.pep

Sequence: 1 VPSYSFIRTAHSEVDLIA 20

Scoring table: PAM 150  
Gap 15

Searched: 179066 seqs, 54579741 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: sptrembl9

1:sp-archaea 2:sp-bacteria 3:sp-fungi 4:sp-human  
5:sp-invertebrate 6:sp-mammal 7:sp-mic 8:sp-organism  
9:sp-phase 10:sp-plant 11:sp-rodent 12:sp-unclassified  
13:sp-vertebrate 14:sp-virus

Statistics: Mean 28.249; Variance 35.475; scale 0.796

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

No.	Score	Query Match	Length	ID	Description	Pred. No.
1	111	82.2	1390	2	069385	GLUCOSYLTRANSFERASE-SI
2	111	82.2	1455	2	069388	GLUCOSYLTRANSFERASE-SI
3	111	82.2	1455	2	069382	GLUCOSYLTRANSFERASE-SI
4	111	82.2	1455	2	069397	GLUCOSYLTRANSFERASE-SI
5	111	82.2	1455	2	069391	GLUCOSYLTRANSFERASE-SI
6	111	82.2	1476	2	069387	GLUCOSYLTRANSFERASE-SI
7	111	82.2	1476	2	069387	GLUCOSYLTRANSFERASE-SI
8	111	82.2	1476	2	069384	GLUCOSYLTRANSFERASE-SI
9	111	82.2	1476	2	069381	GLUCOSYLTRANSFERASE-SI
10	111	82.2	1476	2	069396	GLUCOSYLTRANSFERASE-SI
11	111	82.2	1476	2	069383	GLUCOSYLTRANSFERASE-SI
12	111	82.2	1476	2	069383	GLUCOSYLTRANSFERASE-SI
13	111	82.2	1476	2	069383	GLUCOSYLTRANSFERASE-SI
14	111	82.2	1476	2	069383	GLUCOSYLTRANSFERASE-SI
15	111	82.2	1476	2	069383	GLUCOSYLTRANSFERASE-SI
16	111	82.2	1476	2	069383	GLUCOSYLTRANSFERASE-SI
17	111	82.2	1476	2	069383	GLUCOSYLTRANSFERASE-SI
18	111	82.2	1476	2	069383	GLUCOSYLTRANSFERASE-SI
19	111	82.2	1476	2	069383	GLUCOSYLTRANSFERASE-SI
20	111	82.2	1476	2	069383	GLUCOSYLTRANSFERASE-SI

21	79	58.5	1462	2	069389	GLUCOSYLTRANSFERASE-S
22	78	57.8	575	5	P90900	CYTOLASMIC INTERMEDI
23	77	57.0	1518	2	000600	GLUCOSYLTRANSFERASE I
24	75	55.6	1577	2	054178	GLUCOSYLTRANSFERASE
25	74	54.8	1599	2	000599	GLUCOSYLTRANSFERASE S
26	72	53.3	101	3	006139	CHROMOSOME XII COSMID
27	71	52.6	810	6	077788	NEUROFILAMENT-M SUBUNIT
28	71	52.6	845	11	063370	NEUROFILAMENT PROTEIN
29	71	52.6	1577	2	055265	GLUCOSYLTRANSFERASE PR
30	70	51.9	494	4	016352	NEUROFILAMENT-66.
31	70	51.9	501	11	061958	(CLONE PNF-66).
32	69	51.1	411	13	060057	HYPOTHETICAL 46.2 KD P
33	68	50.4	490	13	P79933	HEPITIN.
34	67	49.6	798	13	090307	HYPOTHETICAL 90.0 KD P
35	66	48.9	913	13	013099	MIDDLE MOLECULAR WEIGH
36	65	48.1	913	13	007962	GELFILTEN.
37	65	48.1	897	13	013098	MIDDLE MOLECULAR WEIGH
38	65	48.1	1110	13	091255	NE-180.
39	64	47.4	342	5	025031	CYSTEINE PROTEINASE.
40	63	46.7	471	13	P87360	HYPOHETICAL 86.2 KD P
41	63	46.7	771	2	P71602	GEFILITIN.
42	62	45.9	132	5	025509	CYTOSOLIC P450 (PRAG
43	62	45.9	265	4	015187	T-CLUSTER BINDING PROT
44	62	45.9	287	2	051618	HYPOTHETICAL 33.6 KD P
45	61	45.2	290	2	083954	P26.

## ALIGNMENTS

RESULT	1	PRELIMINARY:	PRT: 1390 AA.
ID	069385;		
AC	069385;		
DT	01-AUG-1998 (TREMBLREL. 07, CREATED)		
DT	01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)		
DT	01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)		
DE	GLUCOSYLTRANSFERASE-SI.		
GN	GTF.		
OS	STREPTOCOCCUS MUTANS.		
OC	BACTERIA: FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; STREPTOCOCCACEAE;		
OC	STREPTOCOCCUS.		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-MT4245;		
RC	MEDLINE: 98231643.		
RA	FUJIMURA T., TERAO Y., HOSHINO T., KAWABATA S., OOSHIMA T., SOBUE S.,		
RT	"Molecular analyses of glucosyltransferase genes among strains of		
RT	Streptococcus mutans."		
RT	FEMS MICROBIOL. LETT. 161:331-336(1998).		
DR	EMBL: D88655; D1027046; -		
KW	TRANSFERASE.		
SO	SEQUENCE 1390 AA; 155375 MW; C211E7B8 CMC32;		
Query Match			
Best Local Similarity	82.2%; Score 111; DB 2; Length 1390;		
Matches	18; Conservative 0; Mismatches 0; Indels 1; Gaps 1;		
Db	578 VPSYSFIRTAHSEVDLI 595		
Qy	1 VPSYSFIRTAHSEVDLI 19		
RESULT	2	PRELIMINARY:	PRT: 1455 AA.
ID	069388;		
AC	069388;		
DT	01-AUG-1998 (TREMBLREL. 07, CREATED)		
DT	01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)		
DT	01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)		
DE	GLUCOSYLTRANSFERASE-SI.		
GN	GTF.		
OS	STREPTOCOCCUS MUTANS.		
OC	BACTERIA: FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; STREPTOCOCCACEAE;		
OC	STREPTOCOCCUS.		

RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-MT4239;  
 RX MEDLINE: 98231643.  
 RA FUJIMURA T., TERAOKA Y., HOSHINO T., KAWABATA S., OOSHIMA T., SOBUE S.,  
 RA KIMURA S., HAMADA S.,  
 RT "Molecular analyses of glucosyltransferase genes among strains of  
 RT Streptococcus mutans."  
 RL FEMS MICROBIOL. LETT. 161:331-336(1998).  
 DR EMBL: D88658; D1027050;  
 KM TRANSFERASE.  
 SO SEQUENCE 1455 AA; 163046 MW; 7659CPRC CRC32;  
 Query Match 82.2%; Score 111; DB 2; Length 1455;  
 Best Local Similarity 94.7%; Pred. No. 2.83e-12;  
 Matches 18; Conservative 0; Mismatches 0; Indels 1; Gaps 1;  
 DB 578 VPSTSFIR-AHDSEVODLI 595  
 1 VPSTSFIRTAHDSEVODLI 19  
 RESULT 3 PRELIMINARY; PRT; 1455 AA.  
 AC 069382;  
 DT 01-AUG-1998 (TREMBLREL. 07, CREATED)  
 DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
 DE GLUCOSYLTRANSFERASE-SI.  
 GN GTEC.  
 OS STREPTOCOCCUS MUTANS.  
 OC BACTERIA: FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; STREPTOCOCCACEAE;  
 OC STREPTOCOCCUS.  
 [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-MT8148;  
 RX MEDLINE: 98231643.  
 RA FUJIMURA T., TERAOKA Y., HOSHINO T., KAWABATA S., OOSHIMA T., SOBUE S.,  
 RA KIMURA S., HAMADA S.,  
 RT "Molecular analyses of glucosyltransferase genes among strains of  
 RT Streptococcus mutans."  
 RL FEMS MICROBIOL. LETT. 161:331-336(1998).  
 DR EMBL: D88652; D1027042;  
 KM TRANSFERASE.  
 SO SEQUENCE 1455 AA; 162970 MW; 40B022BD CRC32;  
 Query Match 82.2%; Score 111; DB 2; Length 1455;  
 Best Local Similarity 94.7%; Pred. No. 2.83e-12;  
 Matches 18; Conservative 0; Mismatches 0; Indels 1; Gaps 1;  
 DB 578 VPSTSFIR-AHDSEVODLI 595  
 1 VPSTSFIRTAHDSEVODLI 19  
 RESULT 4 PRELIMINARY; PRT; 1455 AA.  
 AC 069397;  
 DT 01-AUG-1998 (TREMBLREL. 07, CREATED)  
 DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
 DE GLUCOSYLTRANSFERASE-SI.  
 GN GTEC.  
 OS STREPTOCOCCUS MUTANS.  
 OC BACTERIA: FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; STREPTOCOCCACEAE;  
 OC STREPTOCOCCUS.  
 [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-MT4467;  
 RX MEDLINE: 98231643.  
 RA FUJIMURA T., TERAOKA Y., HOSHINO T., KAWABATA S., OOSHIMA T., SOBUE S.,  
 RA KIMURA S., HAMADA S.,  
 RT "Molecular analyses of glucosyltransferase genes among strains of

RT Streptococcus mutans."  
 RL FEMS MICROBIOL. LETT. 161:331-336(1998).  
 DR EMBL: D89978; D1027060;  
 KM TRANSFERASE.  
 SO SEQUENCE 1455 AA; 162914 MW; 7D729DA0 CRC32;  
 Query Match 82.2%; Score 111; DB 2; Length 1455;  
 Best Local Similarity 94.7%; Pred. No. 2.83e-12;  
 Matches 18; Conservative 0; Mismatches 0; Indels 1; Gaps 1;  
 DB 578 VPSTSFIR-AHDSEVODLI 595  
 1 VPSTSFIRTAHDSEVODLI 19  
 RESULT 5 PRELIMINARY; PRT; 1455 AA.  
 AC 069391;  
 DT 01-AUG-1998 (TREMBLREL. 07, CREATED)  
 DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
 DE GLUCOSYLTRANSFERASE-SI.  
 GN GTEC.  
 OS STREPTOCOCCUS MUTANS.  
 OC BACTERIA: FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; STREPTOCOCCACEAE;  
 OC STREPTOCOCCUS.  
 [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-MT4251;  
 RX MEDLINE: 98231643.  
 RA FUJIMURA T., TERAOKA Y., HOSHINO T., KAWABATA S., OOSHIMA T., SOBUE S.,  
 RA KIMURA S., HAMADA S.,  
 RT "Molecular analyses of glucosyltransferase genes among strains of  
 RT Streptococcus mutans."  
 RL FEMS MICROBIOL. LETT. 161:331-336(1998).  
 DR EMBL: D88650; D1027053;  
 KM TRANSFERASE.  
 SO SEQUENCE 1476 AA; 165833 MW; 9119C72B CRC32;  
 Query Match 82.2%; Score 111; DB 2; Length 1476;  
 DB 578 VPSTSFIR-AHDSEVODLI 595  
 1 VPSTSFIRTAHDSEVODLI 19  
 RESULT 6 PRELIMINARY; PRT; 1476 AA.  
 AC 069390;  
 DT 01-AUG-1998 (TREMBLREL. 07, CREATED)  
 DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
 DE GLUCOSYLTRANSFERASE-SI.  
 GN GTEB.  
 OS STREPTOCOCCUS MUTANS.  
 OC BACTERIA: FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; STREPTOCOCCACEAE;  
 OC STREPTOCOCCUS.  
 [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-MT4251;  
 RX MEDLINE: 98231643.  
 RA FUJIMURA T., TERAOKA Y., HOSHINO T., KAWABATA S., OOSHIMA T., SOBUE S.,  
 RA KIMURA S., HAMADA S.,  
 RT "Molecular analyses of glucosyltransferase genes among strains of  
 RT Streptococcus mutans."  
 RL FEMS MICROBIOL. LETT. 161:331-336(1998).  
 DR EMBL: D88650; D1027053;  
 KM TRANSFERASE.  
 SO SEQUENCE 1476 AA; 165833 MW; 9119C72B CRC32;  
 Query Match 82.2%; Score 111; DB 2; Length 1476;

Best Local Similarity 94.7%; Pred. No. 2.83e-12;  
Matches 18; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

Db 552 VPSYSFIR-AHSEVODLI 569  
QY 1 VPSYSFIRTAHSEVODLI 19

RESULT 7  
ID 069387 PRELIMINARY; PRT; 1476 AA.  
AC 069387;  
DT 01-AUG-1998 (TREMBLREL. 07, CREATED)  
DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)  
DE 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
GN GTFB.  
OS STREPTOCOCCUS MUTANS.  
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; STREPTOCOCCACEAE;  
OC STREPTOCOCCUS.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-MT4239;  
RX MEDLINE: 98231643.  
RA FUJIMURA T., TERAO Y., HOSHINO T., KAWABATA S., OOSHIMA T., SOBUE S.,  
KIMURA S., HAMADA S.;  
RT "Molecular analyses of glucosyltransferase genes among strains of  
streptococcus mutans.";  
RL FEMS MICROBIOL. LETT. 161:331-336(1998).  
DR EMBL: D88657; D1027049; -.  
KW TRANSFERASE.  
SQ SEQUENCE 1476 AA; 165819 MW; AF6DBA6F CRC32;

Query Match 82.2%; Score 111; DB 2; Length 1476;  
Best Local Similarity 94.7%; Pred. No. 2.83e-12;

Matches 18; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

Db 552 VPSYSFIR-AHSEVODLI 569  
QY 1 VPSYSFIRTAHSEVODLI 19

RESULT 8  
ID 069384 PRELIMINARY; PRT; 1476 AA.  
AC 069384;  
DT 01-AUG-1998 (TREMBLREL. 07, CREATED)  
DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)  
DE 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
GN GTFB.  
OS STREPTOCOCCUS MUTANS.  
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; STREPTOCOCCACEAE;  
OC STREPTOCOCCUS.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-MT4245;  
RX MEDLINE: 98231643.  
RA FUJIMURA T., TERAO Y., HOSHINO T., KAWABATA S., OOSHIMA T., SOBUE S.,  
KIMURA S., HAMADA S.;  
RT "Molecular analyses of glucosyltransferase genes among strains of  
streptococcus mutans.";  
RL FEMS MICROBIOL. LETT. 161:331-336(1998).  
DR EMBL: D88654; D1027045; -.  
KW TRANSFERASE.  
SQ SEQUENCE 1476 AA; 165792 MW; C76D8483 CRC32;

Query Match 82.2%; Score 111; DB 2; Length 1476;  
Best Local Similarity 94.7%; Pred. No. 2.83e-12;

Matches 18; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

Db 552 VPSYSFIR-AHSEVODLI 569  
QY 1 VPSYSFIRTAHSEVODLI 19

RESULT 9  
ID 069381 PRELIMINARY; PRT; 1476 AA.  
AC 069381;  
DT 01-AUG-1998 (TREMBLREL. 07, CREATED)  
DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)  
DE 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
GN GTFB.  
OS STREPTOCOCCUS MUTANS.  
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; STREPTOCOCCACEAE;  
OC STREPTOCOCCUS.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-MT8148;  
RX MEDLINE: 98231643.  
RA FUJIMURA T., TERAO Y., HOSHINO T., KAWABATA S., OOSHIMA T., SOBUE S.,  
KIMURA S., HAMADA S.;  
RT "Molecular analyses of glucosyltransferase genes among strains of  
streptococcus mutans.";  
RL FEMS MICROBIOL. LETT. 161:331-336(1998).  
DR EMBL: D88651; D1027041; -.  
KW TRANSFERASE.  
SQ SEQUENCE 1476 AA; 165638 MW; 0868C20A CRC32;

Query Match 82.2%; Score 111; DB 2; Length 1476;  
Best Local Similarity 94.7%; Pred. No. 2.83e-12;

Matches 18; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

Db 552 VPSYSFIR-AHSEVODLI 569  
QY 1 VPSYSFIRTAHSEVODLI 19

RESULT 10  
ID 069396 PRELIMINARY; PRT; 1476 AA.  
AC 069396;  
DT 01-AUG-1998 (TREMBLREL. 07, CREATED)  
DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)  
DE 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
GN GTFB.  
OS STREPTOCOCCUS MUTANS.  
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; STREPTOCOCCACEAE;  
OC STREPTOCOCCUS.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-MT4467;  
RX MEDLINE: 98231643.  
RA FUJIMURA T., TERAO Y., HOSHINO T., KAWABATA S., OOSHIMA T., SOBUE S.,  
KIMURA S., HAMADA S.;  
RT "Molecular analyses of glucosyltransferase genes among strains of  
streptococcus mutans.";  
RL FEMS MICROBIOL. LETT. 161:331-336(1998).  
DR EMBL: D89977; D1027059; -.  
KW TRANSFERASE.  
SQ SEQUENCE 1476 AA; 165647 MW; 258D4ADD CRC32;

Query Match 82.2%; Score 111; DB 2; Length 1476;  
Best Local Similarity 94.7%; Pred. No. 2.83e-12;

Matches 18; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

Db 552 VPSYSFIR-AHSEVODLI 569  
QY 1 VPSYSFIRTAHSEVODLI 19

RESULT 11  
ID 055263 PRELIMINARY; PRT; 1590 AA.  
AC 055263;  
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)  
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)

DE GLUCOSYLTRANSFERASE GTF-1.  
 OS STREPTOCOCCUS SOBRINUS.  
 OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; STREPTOCOCCACEAE;  
 OC STREPTOCOCCUS.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-ATCC 33478;  
 RA SATO S.;  
 RL ANN. KAGOSHIMA UNIV. DENTAL SCHOOL 16:23-29(1996).  
 DR EMBL: D63570; D1010438;  
 PFAM: PF00128; alpha-amylase; 1.  
 KM TRANSFERASE.  
 SO SEQUENCE 1590 AA; 176057 MW; 4718666A CRC32;  
 Query Match 71.9%; Score 97; DB 2; Length 1590;  
 Best Local Similarity 84.2%; Pred. No. 1.14e-08;  
 Matches 16; Conservative 1; Mismatches 1; Indels 1; Gaps 1;  
 548 VPSYSFIRAHSEVODLI 565  
 :|||||:|||||:|  
 QY 1 VPSYSFIRAHSEVODLI 19

RESULT 12  
 ID Q59983 PRELIMINARY; PRT; 1590 AA.  
 AC Q59983;  
 DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
 DT 01-JUN-1999 (TREMBLREL. 09, LAST ANNOTATION UPDATE)  
 DE GLUCOSYLTRANSFERASE-I PRECURSOR (EC 2.4.1.5) (DEXTRANSUCRASE) (SUCROSE 6-GLUCOSYLTRANSFERASE).  
 GN GTF1.  
 OS STREPTOCOCCUS SOBRINUS.  
 OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; STREPTOCOCCACEAE;  
 OC STREPTOCOCCUS.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-OM2176;  
 RX MEDLINE: 94146405.  
 RA SATO S., INOUE M., HANDA N., AIZAWA Y., ISOBE Y., KATAYAMA T.;  
 RT "DNA sequence of the glucosyltransferase gene of serotype d Streptococcus sobrinus";  
 RT DNA SEQ. 4:19-27(1993).  
 CC -1- CATALYTIC ACTIVITY: SUCROSE + (1,6-ALPHA-D-GLUCOSYL)(N) -  
 CC D-FRUCTOSE + (1,6-ALPHA-D-GLUCOSYL)(N+1).  
 DR EMBL: D13858; D1003482;  
 DB PFAM: PF00128; alpha-amylase; 1.  
 DE SIGNAL; TRANSFERASE; GLUCOSYLTRANSFERASE.  
 GN SIGNAL 1 38 POTENTIAL.  
 FT CHAIN 39 1590 GLUCOSYLTRANSFERASE-I..  
 SQ SEQUENCE 1590 AA; 175955 MW; FE6A57D7 CRC32;  
 Query Match 71.9%; Score 97; DB 2; Length 1590;  
 Best Local Similarity 84.2%; Pred. No. 1.14e-08;  
 Matches 16; Conservative 1; Mismatches 1; Indels 1; Gaps 1;  
 Db 548 VPSYSFIRAHSEVODII 565  
 :|||||:|||||:|  
 QY 1 VPSYSFIRAHSEVODLI 19

RESULT 13  
 ID Q52224 PRELIMINARY; PRT; 1508 AA.  
 AC Q52224;  
 DT 01-JUN-1998 (TREMBLREL. 06, CREATED)  
 DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)  
 DT 01-JUN-1998 (TREMBLREL. 06, LAST ANNOTATION UPDATE)  
 DE GLUCOSYLTRANSFERASE (EC 2.4.1.5) (DEXTRANSUCRASE)  
 DE (SUCROSE 6-GLUCOSYLTRANSFERASE).  
 GN DSRB.  
 OS LEUCONOSTOC MESPENTEROIDES.  
 OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; LACTOBACILLACEAE;  
 OC LEUCONOSTOC.

RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-NRRL B-1299;  
 RA MONCHOIS V., REMAUD-SIMEON M., MONSAN P., WILLENOT R.M.;  
 RL FEMS MICROBIOL. LETT. 0:0-0(1998).  
 CC -1- CATALYTIC ACTIVITY: SUCROSE +  
 CC (1,6-ALPHA-D-GLUCOSYL)(N) - D-FRUCTOSE +  
 CC (1,6-ALPHA-D-GLUCOSYL)(N+1).  
 DR EMBL: AF030129; G2766612;  
 KM TRANSFERASE; GLUCOSYLTRANSFERASE.  
 SO SEQUENCE 1508 AA; 168511 MW; 7D3B6FFA CRC32;  
 Query Match 69.6%; Score 94; DB 2; Length 1508;  
 Best Local Similarity 70.0%; Pred. No. 6.38e-08;  
 Matches 14; Conservative 4; Mismatches 1; Indels 1; Gaps 1;  
 Db 634 IPNYSFVR-AHSEVOTVIA 652  
 :|||||:|||||:|  
 QY 1 VPSYSFIRAHSEVODLIA 20

RESULT 14  
 ID Q48756 PRELIMINARY; PRT; 1290 AA.  
 AC Q48756;  
 DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
 DE DEXTRANSUCRASE.  
 OS LEUCONOSTOC MESPENTEROIDES.  
 OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; LACTOBACILLACEAE;  
 OC LEUCONOSTOC.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-NRRL B1299;  
 RX MEDLINE: 97136686.  
 RA MONCHOIS V., WILLENOT R.M., REMAUD-SIMEON M., CROUX C., MONSAN P.;  
 RT Cloning and sequencing of a gene coding for a novel dextranucrase  
 RT from Leuconostoc mesenteroides NRRL B-1299 synthesizing only alpha  
 RT (1-6) and alpha (1-3) linkages.;  
 RL GENE 182:23-32(1996).  
 DR EMBL: U38181; G1022963;  
 PFAM: PF00128; alpha-amylase; 1.  
 SQ SEQUENCE 1290 AA; 145590 MW; F4460B13 CRC32;  
 Query Match 68.1%; Score 92; DB 2; Length 1290;  
 Best Local Similarity 78.9%; Pred. No. 1.99e-07;  
 Matches 15; Conservative 2; Mismatches 1; Indels 1; Gaps 1;  
 Db 388 PNTSFIR-AHSEVOTIIA 405  
 :|||||:|||||:|  
 QY 2 PNTSFIRAHSEVODLIA 20

RESULT 15  
 ID Q55264 PRELIMINARY; PRT; 1449 AA.  
 AC Q55264;  
 DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
 DE GLUCOSYLTRANSFERASE PRECURSOR.  
 GN GTF1.  
 OS STREPTOCOCCUS SALIVARIUS.  
 OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; STREPTOCOCCACEAE;  
 OC STREPTOCOCCUS.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 95122197.  
 RA SIMPSON C.L., GIFFARD P.M., JACQUES N.A.;  
 RT "Streptococcus salivarius ATCC 25975 possesses at least two genes  
 RT coding for primer-independent glucosyltransferases";  
 RL INFECT. IMMUN. 63:609-621(1995).  
 DR EMBL: U35495; G662379;  
 PFAM: PF00128; alpha-amylase; 1.

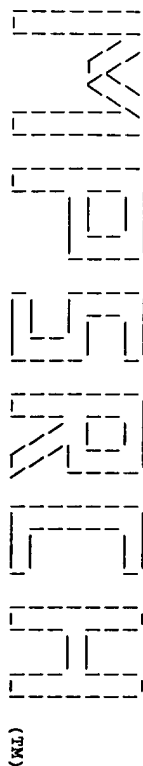
KM SIGNAL: TRANSFERASE. 1 35 POTENTIAL.  
FT SIGNAL 1 36 1449 GLUCOSYLTRANSFERASE.  
FT CHAIN 36 1449  
SQ SEQUENCE 1449 AA; 159984 MW; E9A4BA87 CRC32;

Query Match 60.0%; Score 81; DB 2; Length 1449;  
Best Local Similarity 55.0%; Pred. No. 8.67e-05;  
Matches 11; Conservative 7; Mismatches 1; Indels 1; Gaps 1;

Db 607 MANYAFVR-AHDSEVQSIIG 625  
QY 1 VPSYSFIRRAHDSSEVQDLIA 20

Search completed: Tue Jan 11 15:45:45 2000  
Job time : 47 secs.





Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MSrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Tue Jan 11 15:47:50 2000; Maspar time 2.45 seconds

Target output not generated. 254.053 Million cell updates/sec

File: >US-09-290-049-3

Description: (1-22) from US9290049.pep

Sequence: 1 TGARTINGQLLYFRANGVQVG 22

Scoring table: PAM 150  
Gap 15

Searched: 77977 seqs, 28268293 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: swiss-prot37  
1:swissprot

Statistics: Mean 29.681; Variance 38.978; scale 0.761

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
1	155	100.0	1475	1	GTFF_STRMU	GLUCOSYLTRANSFERASE-I	9.19e-23
2	114	73.5	1597	1	GTFF_STRDO	GLUCOSYLTRANSFERASE-I	4.76e-12
3	112	72.3	1592	1	GTFF_STRDO	GLUCOSYLTRANSFERASE-I	1.49e-11
4	106	68.4	1375	1	GTFF_STRMU	GLUCOSYLTRANSFERASE-I	4.46e-10
5	104	67.1	1365	1	GTFF_STRDO	GLUCOSYLTRANSFERASE-I	1.36e-09
6	93	60.0	1430	1	GTFF_STRMU	GLUCOSYLTRANSFERASE-S	5.53e-07
7	79	51.0	382	1	YJOT_YEAST	HYPOTHETICAL 44.9 KD P	7.50e-04
8	72	46.5	126	1	DOC_BPPI	DEATH ON CURING PROTEIN	2.20e-02
9	70	45.2	529	1	YJ19_NYOP	HYPOTHETICAL 59.0 KD P	5.58e-02
10	67	43.2	2710	1	TOXA_CDADI	TOXIN A	2.19e-01
11	65	41.9	349	1	PTER_MOUSE	PHOSPHOTRIESTERASE REL	5.32e-01
12	65	41.9	1396	1	VLFE_BPTS	L-SHAPED TAIL FIBRE PR	5.32e-01
13	63	40.6	301	1	LECI_MOUSE	ASIALOGLYCOPROTEIN REC	1.27e+00
14	63	40.6	310	1	YC08_PYRHO	HYPOTHETICAL SUGAR KIN	1.27e+00
15	63	40.6	349	1	PTER_RAT	PHOSPHOTRIESTERASE REL	1.27e+00
16	63	40.6	396	1	DUT_VZVD	DEOXYRIBIDINE 5'-TRIPHOS	1.27e+00
17	63	40.6	404	1	CAGS_CHICK	ALPHA-N-ACETYLDIALACTOS	1.27e+00
18	63	40.6	475	1	AP54_YEAST	CLATHRIN COAT ASSEMBLY	1.27e+00
19	63	40.6	1082	1	RPO_ROTPC	RNA-DIRECTED RNA POLYM	1.27e+00
20	62	40.0	464	1	VL2_HPV27	MINOR CAPSID PROTEIN L	1.95e+00
21	62	40.0	467	1	ISP6_SCHPO	SEXUAL DIFFERENTIATION	1.95e+00
22	62	40.0	524	1	VL2_HPV2A	MINOR CAPSID PROTEIN L	1.95e+00
23	39.4	39.4	336	1	VINT_BP186	INTEGRASE	2.97e+00

RESULT ID	1	STANDARD:	PRT:	1475 AA.
AC	GTFF_STRMU			
AD	P08987:			
DT	01-NOV-1988 (REL. 09, CREATED)			
DT	01-NOV-1988 (REL. 09, LAST SEQUENCE UPDATE)			
DT	15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)			
DE	GLUCOSYLTRANSFERASE-I PRECURSOR (EC 2.4.1.5) (GTFF-I) (DEXTRANSUCRASE)			
DE	(SDROSE 6-GLUCOSYLTRANSFERASE).			
GN	GTFF.			
OS	STREPTOCOCCUS MUTANS.			
OC	BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; STREPTOCOCCACEAE;			
OC	STREPTOCOCCUS.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-GS-5;			
RX	MEDLINE: 87306013.			
RA	SHIMOZA T., UEDA S., KURAMITSU H.K.;			
RT	*Sequence analysis of the gtff gene from Streptococcus mutans.*;			
RL	J. BACTERIOL. 169:4263-4270(1987).			
CC	- FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS, THAT ARE THOUGHT TO PLAY A KEY ROLE IN THE DEVELOPMENT OF THE DENTAL PLAQUE BECAUSE OF THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDIANE THE AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS.			
CC	- CATALYTIC ACTIVITY: SUCROSE + (1,6-ALPHA-D-GLUCOSYL) (N+1).			
CC	- SUBCELLULAR LOCATION: SECRETED.			
CC	- DISEASE: DENTAL CARIES.			
CC	- D-PRUCOSE + (1,6-ALPHA-D-GLUCOSYL) (N+1).			
CC	- GTFF-I SYNTHESIZES WATER-INSOLUBLE GLUCANS (ALPHA 1,3-LINKED GLUCOSE AND SOME 1,6 LINKAGES), GTFF-S SYNTHESIZES WATER-SOLUBLE GLUCANS (ALPHA 1,6-GLUCOSE). GTFF-SI SYNTHESIZES BOTH FORMS OF GLUCANS.			
CC	- SIMILARITY: TO OTHER GLUCOSYLTRANSFERASES AND SOME TO A GLUCAN-BINDING PROTEIN FROM S. MUTANS.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> or send an email to <a href="mailto:license@sib-sib.ch">license@sib-sib.ch</a> ).			
CC	-----			
DR	EMBL: M17361; G153640; -			
DR	PIR: B33135; B33135.			
DR	PIR: PF00128; alpha-amylase; 1.			
DR	TRANSFERASE; GLUCOSYLTRANSFERASE; SIGNAL; REPEAT; DENTAL CARIES.			
FT	SIGNAL	1	34	POTENTIAL.

FT CHAIN 35 1475 GLUCOSYLTRANSFERASE-I.  
 FT DOMAIN 35 1050 CATALYTIC (APPROXIMATE).  
 FT REPEAT 1096 1475 GLUCAN-BINDING (APPROXIMATE).  
 FT REPEAT 1096 1129 A REPEAT.  
 FT DOMAIN 1160 1469 5 X TANDEM REPEATS.  
 FT REPEAT 1160 1209 1.  
 FT REPEAT 1224 1274 2.  
 FT REPEAT 1289 1339 3.  
 FT REPEAT 1354 1404 4.  
 FT REPEAT 1419 1469 5.  
 SO SEQUENCE 1475 AA; 165812 MW; 4542CID6 CRC32;

Query Match 100.0%; Score 155; DB 1; Length 1475;  
 Best Local Similarity 100.0%; Pred. No. 9,19e-23;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1300 TGARTINGOLLYFRANGOVKG 1321  
 1 TGARTINGOLLYFRANGOVKG 22

RESULT 2  
 ID GTF2\_STRDO STANDARD; PRT; 1597 AA.  
 AC P11001;  
 DT 01-JUL-1989 (REL. 11, CREATED)  
 DT 01-JUL-1989 (REL. 11, LAST SEQUENCE UPDATE)  
 DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)  
 DE GLUCOSYLTRANSFERASE-1 PRECURSOR (EC 2.4.1.5) (GTF-I) (DEXTRANSUCRASE)  
 DE (SUCROSE 6-GLUCOSYLTRANSFERASE).  
 GN GTFI.  
 OS STREPTOCOCCUS DORNI (STREPTOCOCCUS SOBRINUS).  
 OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; STREPTOCOCCACEAE;  
 OC STREPTOCOCCUS.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-MEE28;  
 RX MEDLINE; 87308014.  
 RA FERRETTI J.J., GILPIN M.L., RUSSELL R.R.B.;  
 RT "Nucleotide sequence of a glucosyltransferase gene from Streptococcus  
 sobrinus MEE28.";  
 RL J. BACTERIOL. 169:4271-4278(1987).  
 CC -1- FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS, THAT ARE THOUGHT  
 TO PLAY A KEY ROLE IN THE DEVELOPMENT OF THE DENTAL PLAQUE BECAUSE  
 OF THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDIATE THE  
 AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS.  
 CC -1- CATALYTIC ACTIVITY: SUCROSE + (1,6-ALPHA-D-GLUCOSYL)(N) -  
 D-FRUCTOSE + (1,6-ALPHA-D-GLUCOSYL)(N+1).  
 CC -1- SUBCELLULAR LOCATION: SECRETED.  
 CC -1- DISEASE: DENTAL CARIES.  
 CC -1- GTF-I SYNTHESIZES WATER-INSOLUBLE GLUCANS (ALPHA 1,3-LINKED  
 GLUCOSE AND SOME 1,6 LINKAGES), GTF-S SYNTHESIZES WATER-SOLUBLE  
 GLUCANS (ALPHA 1,6-GLUCOSE). GTF-SI SYNTHESIZES BOTH FORMS OF  
 GLUCANS.  
 CC -1- SIMILARITY: TO OTHER GLUCOSYLTRANSFERASES AND SOME TO A GLUCAN-  
 BINDING PROTEIN FROM S.MUTANS.  
 CC -1- This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sid.ch](mailto:license@isb-sib.ch)).  
 CC EMBL; M17391; G153647; -  
 DR PIR; P00128; alpha-amylase; 1.  
 DR PFAM; P00128; alpha-amylase; 1.  
 KW TRANSFERASE; GLYCOSYLTRANSFERASE; SIGNAL; REPEAT; DENTAL CARIES.  
 FT SIGNAL 1 38  
 FT CHAIN 39 1597 POTENTIAL.  
 FT DOMAIN 39 1050 GLUCOSYLTRANSFERASE-I.  
 FT DOMAIN 1099 1597 CATALYTIC (APPROXIMATE).  
 FT DOMAIN 1099 1597 GLUCAN-BINDING (APPROXIMATE).  
 FT REPEAT 1099 1132 1.25 A, 2 B, AND 5 AC REPEATS.  
 FT REPEAT 1099 1132 A REPEAT.

FT REPEAT 1163 1213 AC REPEAT.  
 FT REPEAT 1227 1277 AC REPEAT.  
 FT REPEAT 1292 1342 AC REPEAT.  
 FT REPEAT 1352 1399 B REPEAT.  
 FT REPEAT 1406 1455 AC REPEAT.  
 FT REPEAT 1465 1512 B REPEAT.  
 FT REPEAT 1519 1568 AC REPEAT.  
 FT REPEAT 1582 1597 A REPEAT (INCOMPLETE).  
 SO SEQUENCE 1597 AA; 177080 MW; 995233CA CRC32;

Query Match 73.5%; Score 114; DB 1; Length 1597;  
 Best Local Similarity 77.3%; Pred. No. 4,74e-12;  
 Matches 17; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Db 1303 TGAOTIKGOLLYFRANGOVKG 1324  
 1 TGARTINGOLLYFRANGOVKG 22

RESULT 3  
 ID GTF2\_STRDO STANDARD; PRT; 1592 AA.  
 AC P27470;  
 DT 01-AUG-1992 (REL. 23, CREATED)  
 DT 01-AUG-1992 (REL. 23, LAST SEQUENCE UPDATE)  
 DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)  
 DE GLUCOSYLTRANSFERASE-1 PRECURSOR (EC 2.4.1.5) (GTF-I) (DEXTRANSUCRASE)  
 DE (SUCROSE 6-GLUCOSYLTRANSFERASE).  
 GN STREPTOCOCCUS DORNI (STREPTOCOCCUS SOBRINUS).  
 OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; STREPTOCOCCACEAE;  
 OC STREPTOCOCCUS.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-6715;  
 RX MEDLINE; 91123227.  
 RA ABO H., MATSUMURA T., KODAMA T., OHTA H., FUKUI K., KATO K.,  
 RA KAGAWA H.;  
 RT "Peptide sequences for sucrose splitting and glucan binding within  
 Streptococcus sobrinus glucosyltransferase (water-insoluble glucan  
 synthetase).";  
 RL J. BACTERIOL. 173:989-996(1991).  
 CC -1- FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS, THAT ARE THOUGHT  
 TO PLAY A KEY ROLE IN THE DEVELOPMENT OF THE DENTAL PLAQUE BECAUSE  
 OF THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDIATE THE  
 AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS.  
 CC -1- CATALYTIC ACTIVITY: SUCROSE + (1,6-ALPHA-D-GLUCOSYL)(N) -  
 D-FRUCTOSE + (1,6-ALPHA-D-GLUCOSYL)(N+1).  
 CC -1- SUBCELLULAR LOCATION: SECRETED.  
 CC -1- DISEASE: DENTAL CARIES.  
 CC -1- GTF-I SYNTHESIZES WATER-INSOLUBLE GLUCANS (ALPHA 1,3-LINKED  
 GLUCOSE AND SOME 1,6 LINKAGES), GTF-S SYNTHESIZES WATER-SOLUBLE  
 GLUCANS (ALPHA 1,6-GLUCOSE). GTF-SI SYNTHESIZES BOTH FORMS OF  
 GLUCANS.  
 CC -1- SIMILARITY: TO OTHER GLUCOSYLTRANSFERASES AND SOME TO A GLUCAN-  
 BINDING PROTEIN FROM S.MUTANS.  
 CC -1- This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC EMBL; D90213; G217033; -  
 DR PIR; A38175; A38175.  
 DR PFAM; P00128; alpha-amylase; 1.  
 DR HSP; P00655; 2HEB.  
 KW TRANSFERASE; GLYCOSYLTRANSFERASE; SIGNAL; REPEAT; DENTAL CARIES.  
 FT SIGNAL 1 38  
 FT CHAIN 39 1592 POTENTIAL.  
 FT DOMAIN 39 1044 GLUCOSYLTRANSFERASE-I.  
 FT DOMAIN 1093 1592 CATALYTIC (APPROXIMATE).  
 FT DOMAIN 1093 1592 GLUCAN-BINDING (APPROXIMATE).  
 FT REPEAT 1093 1592 6.5 X TANDEM REPEATS.



```

FT REPEAT 1093 1142 1.
FT REPEAT 1158 1207 2.
FT REPEAT 1222 1272 3.
FT REPEAT 1287 1337 4.
FT REPEAT 1402 1451 5.
FT REPEAT 1514 1563 6.
FT REPEAT 1577 1592 7. (INCOMPLETE).
SQ SEQUENCE 1592 AA; 176167 MW; 2734237A CRC32;

Query Match 72.3%; Score 112; DB 1; Length 1592;
Best Local Similarity 77.3%; Pred. No. 1,49e-11;
Matches 17; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Db 1298 TGAORINGOKLYFRANGGOVKG 1319
OY 1 TGAORINGOKLYFRANGGOVKG 22

RESULT 4
ID GTFC_STRMG STANDARD: PRT: 1375 AA.
AC P13470; P05427;
DT 01-NOV-1988 (REL. 09, CREATED)
DT 01-JAN-1990 (REL. 13, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE GLUCOSYLTRANSFERASE-SI PRECURSOR (EC 2.4.1.5) (GTF-SI)
DE (DEXTRANSUCRASE) (SUCROSE 6-GLUCOSYLTRANSFERASE).
GN GTFC.
OS STREPTOCOCCUS MUTANS.
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; STREPTOCOCCACEAE;
CC STREPTOCOCCUS.
CC [1]
RP SEQUENCE FROM N.A.
RC STRAIN-GS-5;
RX MEDLINE: 89137980.
RA UEDA S., SHIROZA T., KURAMITSU H.K.;
RT "Sequence analysis of the gtfc gene from Streptococcus mutans GS-5.";
RL GENE 69:101-109(1988).
RN [2]
RP SEQUENCE OF 1-349 FROM N.A.
RC STRAIN-GS-5;
RX MEDLINE: 87308013.
RA SHIROZA T., UEDA S., KURAMITSU H.K.;
RT "Sequence analysis of the gtfr gene from Streptococcus mutans.";
RL J. BACTERIOL. 169:4263-4270(1987).
CC -1- FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS, THAT ARE THOUGHT
CC TO PLAY A KEY ROLE IN THE DEVELOPMENT OF THE DENTAL PLAQUE BECAUSE
CC OF THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDIATE THE
CC AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS.
CC -1- CATALYTIC ACTIVITY: SUCROSE + (1,6-ALPHA-D-GLUCOSYL) (N) =
CC D-FRUCTOSE + (1,6-ALPHA-D-GLUCOSYL) (N+1).
CC -1- SUBCELLULAR LOCATION: SECRETED.
CC -1- DISEASE: DENTAL CARIES.
CC -1- GTF-I SYNTHESIZES WATER-INSOLUBLE GLUCANS (ALPHA 1,3-LINKED
CC GLUCOSE AND SOME 1,6 LINKAGES). GTF-S SYNTHESIZES WATER-SOLUBLE
CC GLUCANS (ALPHA 1,6-GLUCOSE). GTF-SI SYNTHESIZES BOTH FORMS OF
CC GLUCANS.
CC -1- SIMILARITY: TO OTHER GLUCOSYLTRANSFERASES AND SOME TO A GLUCAN-
CC BINDING PROTEIN FROM S. MUTANS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/)
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M2054; G153643; -
DR EMBL: M17361; G153641; -
DR PIR: J70345; J70345.
DR PIR: C3135; C3135.
DR PFAM: PF00128; alpha-amylase; 1.
DR TRANSFAB: GLYCOSYLTRANSFERASE; SIGNAL; REPEAT; DENTAL CARIES.
KM

```

```

FT SIGNAL 1 34 GLUCOSYLTRANSFERASE-ST.
FT CHAIN 35 1375 CATALYTIC (APPROXIMATE).
FT DOMAIN 35 1050
FT DOMAIN 1126 1375 GLUCAN-BINDING (APPROXIMATE).
FT DOMAIN 1126 1375 2.4 A, 1 C AND 1 AC REPEATS.
FT REPEAT 1126 1159 A REPEAT.
FT REPEAT 1169 1200 A REPEAT.
FT REPEAT 1227 1238 C REPEAT.
FT REPEAT 1253 1303 AC REPEAT.
FT REPEAT 1318 1330 A REPEAT (INCOMPLETE).
SQ SEQUENCE 1375 AA; 153022 MW; 3EA3727E CRC32;

Query March 68.4%; Score 106; DB 1; Length 1375;
Best Local Similarity 68.2%; Pred. No. 4,46e-10;
Matches 15; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

Db 1264 TGTVTNGCARLYFKFNGVQANG 1285
Cy 1 TGAATNGQLLYFRANGVQYKNG 22
||:|:|||||:|||||
|:|:|:|:|:|:|:|:|

RESULT 5
ID GTF_S PRDNO STANDARD; PRT; 1365 AA.
AC P29336;
DT 01-DEC-1992 (REL. 24, CREATED)
DT 01-DEC-1992 (REL. 24, LAST SEQUENCE UPDATE)
DT 01-APR-1993 (REL. 25, LAST ANNOTATION UPDATE)
DE GLUCOSYLTRANSFERASE-S PRECURSOR (EC 2.4.1.5) (GTF-S) (DEXTRANSUCRASE)
DE (SUCROSE 6-GLUCOSYLTRANSFERASE).
GN GTF_S.
OS STREPTOCOCCUS DOWNEI (STREPTOCOCCUS SOBRIUS).
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; STREPTOCOCCACEAE;
OC STREPTOCOCCUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-MF28;
RX MEDLINE; 90316665.
RA GILMORE K.S., RUSSELL R.R., FERRETTI J.J.;
RT "Analysis of the Streptococcus downei gtf gene, which specifies a
RT glucosyltransferase that synthesizes soluble glucans.";
RL INFECT. IMMUN. 58:2452-2458(1990).
CC -1- FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS, THAT ARE THOUGHT
CC TO PLAY A KEY ROLE IN THE DEVELOPMENT OF THE DENTAL PLAQUE BECAUSE
CC OF THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDIANE THE
CC AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS.
CC -1- CATALYTIC ACTIVITY: SUCROSE + (1,6-ALPHA-D-GLUCOSYL)(N) =
CC D-FRUCTOSE + (1,6-ALPHA-D-GLUCOSYL)(N+1).
CC -1- ENZYME REGULATION: GLUCAN SYNTHESIS BY GTF-S IS INDEPENDENT OF
CC PRIMER GLUCAN UNLIKE GTF-I.
CC -1- DISEASE: DENTAL CARIES.
CC -1- GTF-S SYNTHESIZES WATER-SOLUBLE GLUCANS (ALPHA 1,6-GLUCOSE).
CC -1- SIMILARITY: TO OTHER GLUCOSYLTRANSFERASES AND SOME TO A GLUCAN-
CC BINDING PROTEIN FROM S. MUTANS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M30943; G153653; -.
CC PIR; A41483; A41483.
CC PFAM; PF00128; alpha-amylase; 1.
CC TRANSFRASE; GLYCOSYLTRANSFERASE; SIGNAL; REPEAT; DENTAL CARIES.
CC SIGNAL 1 36 OR 37 (POTENTIAL).
CC CHAIN 37 1365 GLUCOSYLTRANSFERASE-S.
CC DOMAIN 37 1050 CATALYTIC (APPROXIMATE).
CC DOMAIN 1083 1365 GLUCAN-BINDING (APPROXIMATE).
CC DOMAIN 1083 1365 4.5 X TANDEM REPEATS.
CC REPEAT 1083 1131 1.
CC REPEAT 1150 1199 2.

```

FT REPEAT 1225 1274 3.  
 FT REPEAT 1289 1339 4.  
 FT REPEAT 1353 1365 5. (INCOMPLETE).  
 SQ SEQUENCE 1365 AA; 151590 MM; 156F8B31 CRC32;

Query Match 67.1% Score 104; DB 1; Length 1365;  
 Best Local Similarity 63.6% Pred. No. 1.35e-09;  
 Matches 14; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

DB 1235 TGEOTIDGKVFQDNGVQVK 1256  
 1 TGAATNGQLLYFRANGVQVK 22

RESULT 6  
 ID GTFD-STRMU STANDARD; PRT; 1430 AA.  
 AC P49331;  
 DT 01-FEB-1996 (REL. 33, CREATED)  
 DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)  
 DE 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
 DE GLUCOSYLTRANSFERASE-S PRECURSOR (EC 2.4.1.5) (GTF-S) (DEXTRANSUCRASE)  
 DE (SUCCOSE 6-GLUCOSYLTRANSFERASE).  
 GN GTFD.  
 OS STREPTOCOCCUS MUTANS.  
 OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; STREPTOCOCCACEAE;  
 OC STREPTOCOCCUS.  
 RN 1)  
 RP SEQUENCE FROM N.A.  
 RX STRAIN-GS-5;  
 RX MEDLINE: 91100958.  
 RA HONDA O., KATO C., KUDAMITSU H.K.;  
 RT "Nucleotide sequence of the Streptococcus mutans gtfD gene encoding  
 the glucosyltransferase-S enzyme."  
 RL J. GEN. MICROBIOL. 136:2099-2105(1990).  
 CC -1- FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS, THAT ARE THOUGHT  
 TO PLAY A KEY ROLE IN THE DEVELOPMENT OF THE DENTAL PLAQUE BECAUSE  
 OF THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDIATE THE  
 AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS.  
 CC -1- CATALYTIC ACTIVITY: SUCROSE + (1,6-ALPHA-D-GLUCOSYL)(N) -  
 CC -1- D-FRUCTOSE + (1,6-ALPHA-D-GLUCOSYL)(N+1).  
 CC -1- SUBCELLULAR LOCATION: SECRETED.  
 CC -1- DISEASE: DENTAL CARIES.  
 CC -1- GTF-1 SYNTHESIZES WATER-INSOLUBLE GLUCANS (ALPHA 1,3-LINKED  
 GLUCOSE AND SOME 1,6 LINKAGES), GTF-S SYNTHESIZES WATER-SOLUBLE  
 GLUCANS (ALPHA 1,6-GLUCOSE). GTF-S1 SYNTHESIZES BOTH FORMS OF  
 GLUCANS.  
 CC -1- SIMILARITY: TO OTHER GLUCOSYLTRANSFERASES AND SOME TO A GLUCAN-  
 BINDING PROTEIN FROM S. MUTANS.  
 CC -1- This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC EMBL: M29296; G153645; -  
 DR PFAM: PF00128; alpha-amylase; 1.  
 KM TRANSFERASE: GLUCOSYLTRANSFERASE: SIGNAL: REPEAT: DENTAL CARIES.  
 FT SIGNAL 1  
 FT CHAIN 1 1430  
 FT DOMAIN 1232 1423 3 X 63 AA APPROXIMATE TANDEM REPEATS.  
 FT REPEAT 1232 1295 1.  
 FT REPEAT 1296 1359 2.  
 FT REPEAT 1360 1423 3.  
 SO SEQUENCE 1430 AA; 159765 MM; D3DE3681 CRC32;

Query Match 60.0% Score 93; DB 1; Length 1430;  
 Best Local Similarity 59.1% Pred. No. 5.53e-07;  
 Matches 13; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 1 TGAATNGQLLYFRANGVQVK 22  
 4 RTINGQLLYFRANGVQVK 22

RESULT 7  
 ID YJ07\_YEAST STANDARD; PRT; 382 AA.  
 AC P47007;  
 DT 01-FEB-1996 (REL. 33, CREATED)  
 DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)  
 DE 01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)  
 DE HYPOTHETICAL 44.9 KD PROTEIN IN INO1-1DS2 INTERGENIC REGION.  
 GN YJ147C OR J0639.  
 OS SACCAROMYCES CEREVISIAE (BAKER'S YEAST).  
 OC EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOMYCETES; SACCAROMYCETALES;  
 OC SACCAROMYCETACEAE; SACCAROMYCES.  
 RN 1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-S288C / FY1679;  
 RX MEDLINE: 96408771.  
 RA KATSIOLOU C., TZEREMIA M., TAVERNARAKIS N., ALEXANDRAKI D.;  
 RT "Sequence analysis of a 40.7 kb segment from the left arm of yeast  
 RT chromosome X reveals 14 known genes and 13 new open reading frames  
 RT including homologues of genes clustered on the right arm of  
 RT chromosome XI."  
 RL YEAST 12:787-797(1996).  
 CC -1- This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC EMBL: Z49422; G1015561; -  
 DR EMBL: X67371; G854550; -  
 KM HYPOTHETICAL PROTEIN.  
 SQ SEQUENCE 382 AA; 44862 MM; 5B128AB9 CRC32;

Query Match 51.0% Score 79; DB 1; Length 382;  
 Best Local Similarity 38.9% Pred. No. 7.50e-04;  
 Matches 7; Conservative 7; Mismatches 4; Indels 0; Gaps 0;

DB 170 RVASITLYRTHDVQIR 187  
 4 RTINGQLLYFRANGVQVK 21

RESULT 8  
 ID DOC\_BPPI STANDARD; PRT; 126 AA.  
 AC Q06259;  
 DT 01-FEB-1995 (REL. 31, CREATED)  
 DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)  
 DE DEATH ON CURING PROTEIN.  
 GN DOC.  
 OS BACTERIOPHAGE P1.  
 OC VIRUSES; DSDNA VIRUSES, NO RNA STAGE; TAILED PHAGES; MYOVIRIDAE.  
 RN 1)  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 94016561.  
 RA LEHNHERR H., MAGUIN E., JAFRI S., YARMOLINSKY M.B.;  
 RT "Plasmid addiction genes of bacteriophage P1: doc, which causes cell  
 RT death on curing of prophage, and phd, which prevents host death when  
 RT prophage is retained."  
 RL J. MOL. BIOL. 233:414-428(1993).  
 CC -1- FUNCTION: KILLS CELLS. DOC AND PHD PROTEIN FUNCTION IN UNISSON TO  
 CC STABILIZE PLASMID NUMBER BY INDUCING A LETHAL RESPONSE TO PLASMID  
 CC LOSS.  
 CC -1- THE CONCENTRATION OF PHD IN P1 LYSOGENS IS FAR GREATER THAN THAT  
 CC OF THE POISON IT ANTAGONIZES. SUCH AN EXCESS MAY ASSUME THE  
 CC WELL-BEING OF CARRIERS OF THE ADDICTING PLASMID.

CC BACTERIA/FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP: CLOSTRIDIACEAE;  
CC CLOSTRIDIUM.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-VPI 10463;  
RX MEDLINE: 90221894.  
RA SAVERBORN M., VON EICHEL-STREIBER C.;  
RT "Nucleotide sequence of Clostridium difficile toxin A.";  
RL NUCLEIC ACIDS RES. 18:1629-1630(1990).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-VPI 10463;  
RX MEDLINE: 90129305.  
RA DOVE C.H., WANG S.Z., PRICE S.B., PHELPS C.J., LYERLY D.M.,  
RT WILKINS T.W., JOHNSON J.L.;  
RL "Molecular characterization of the Clostridium difficile toxin A  
gene.";  
RN INFECT. IMMUN. 58:480-488(1990).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN-VPI 10463;  
RA VON EICHEL-STREIBER C.;  
RT SUBMITTED (JAN-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
RL -1 DOMAIN: THE C-TERMINAL PART OF TOXIN A CONSISTS OF A 833 AA  
CC REPETITIVE STRUCTURE. THIS PART OF TOXIN A IS COMPOSED OF FIVE  
CC DIFFERENT OLIGOPEPTIDES.  
CC -1 DISEASE: CLOSTRIDIUM DIFFICILE PRODUCES TWO TOXINS, AN  
CC ENTEROTOXIN CALLED A AND CYTOTOXIN B. ONLY AFTER THE ENTERAL  
CC DELIVERY OF THE ENTEROTOXIN A MAY THE CHARACTERISTIC DISEASE  
CC CALLED PSEUDOMEMBRANOUS COLITIS BE INDUCED.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>  
CC or send an email to [license@isb-sdb.ch](mailto:license@isb-sdb.ch)).  
CC -----  
DR EMBL: X51797; G40441; -;  
DR EMBL: M30307; G144926; -;  
DR EMBL: X92982; E212010; -;  
DR PIR: S08638; S08638.  
KW TOXIN; ENTEROTOXIN.  
SQ SEQUENCE 2710 AA; 308052 MW; D04C4C08 CRC32;  
  
Query Match 43.2%; Score 67; DB 1; Length 2710;  
Best Local Similarity 52.2%; Pred. No. 2.19e-01;  
Matches 12; Conservative 3; Mismatches 7; Indels 1; Gaps 1;  
  
Db 2460 TGLRTIDGKKRYTNTAVAVTG 2482  
|||::|||::|  
QY 1 TGARTINGQLLYFRAN-GVGYKG 22  
  
RESULT 11  
ID PTER\_MOUSE STANDARD; PRT; 349 AA.  
AC 060866;  
DT 15-DEC-1998 (REL. 37, CREATED)  
DT 15-DEC-1998 (REL. 37, LAST SEQUENCE UPDATE)  
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)  
DE PROSPTHESTERASE RELATED PROTEIN (PARATHION HYDROLASE-RELATED  
DE PROTEIN).  
DE PTER.  
GN MUS MUSCULUS (MOUSE).  
OS EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;  
OC RODENTIA; SCUROGNATHI; MORIDAE; MORINAE; MUS.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-C57BL/6J; TISSUE-KIDNEY;  
RX MEDLINE: 96194889.  
RA HOU X., MASER R.L., MAGENHEIMER B.S., CALVET J.P.;  
RT "A mouse kidney- and liver-expressed cDNA having homology with a

RT PROKARYOTIC PARATHYROID HYDROLASE (PHOSPHOTRIESTERASE)-encoding gene:  
 RT abnormal expression in injured and polycystic kidneys.  
 RL GENE 168:157-163(1996).  
 CC -1- COFACTOR: CONTAINS 2 MOLES OF ZINC PER SUBUNIT (BY SIMILARITY).  
 CC -1- TISSUE SPECIFICITY: EXPRESSED IN KIDNEY AND LIVER.  
 CC -1- SIMILARITY: BELONGS TO THE PHOSPHOTRIESTERASE FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: U28016; G881499. -  
 DR MGD: MGI:107372; PTER.  
 KW HYDROLASE; ZINC.  
 KM METAL 26 26 ZINC 1 (BY SIMILARITY).  
 FT METAL 26 26 ZINC 1 (BY SIMILARITY).  
 FT METAL 169 169 ZINC 1 AND 2 (BY SIMILARITY).  
 FT METAL 201 201 ZINC 2 (BY SIMILARITY).  
 FT METAL 230 230 ZINC 2 (BY SIMILARITY).  
 FT METAL 298 298 ZINC 1 (BY SIMILARITY).  
 SQ SEQUENCE 349 AA; 39218 MW; 58404EEF CMC32.  
 Query Match 41.9%; Score 65; DB 1; Length 349;  
 Best Local Similarity 53.8%; Pred. No. 5.32e-01;  
 Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;  
 DB 79 AIRELLIYKAKG 91  
 QY 1:|||||:1:1  
 5 TINGOLLYFRANG 17  
 RESULT 12  
 ID VITE.BPTS STANDARD; PRT; 1396 AA.  
 AC P13350; Q48502.  
 DT 01-JAN-1990 (REL. 13, CREATED)  
 DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)  
 DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)  
 DE L-SHAPED TAIL FIBRE PROTEIN (LTF PROTEIN).  
 GN LTF  
 OS BACTERIOPHAGE T5.  
 OC VIRUSES; DSDNA VIRUSES, NO RNA STAGE; TAILED PHAGES; SIPHOVIRIDAE.  
 RN 11  
 RP SEQUENCE FROM N.A.  
 RM MEDLINE: 95309401.  
 RA KALIMAN A.V., KULSHIN V.E., SHLYAPNIKOV M.G., KSENZENKO V.N.,  
 RA KRYUKOV V.M.;  
 RT The nucleotide sequence of the bacteriophage T5 ltf gene.\*;  
 RL FEBS LETT. 366:46-48(1995).  
 CC [2]  
 CC SEQUENCE FROM N.A.  
 RA KALIMAN A.V.;  
 RP SUBMITTED (JAN-1998) TO EMBL/GENBANK/DBJ DATA BANKS.  
 RN 13  
 RP PRELIMINARY PARTIAL SEQUENCE FROM N.A.  
 RX MEDLINE: 88289370.  
 RA KALIMAN A.V., KRYUKOV V.M., BAYEV A.A.;  
 RT The nucleotide sequence of bacteriophage T5 DNA at the region  
 RT between early and late genes.\*;  
 RL NUCLEIC ACIDS RES. 16:6230-6230(1988).  
 CC -1- FUNCTION: NONSENSE/PROTEIN THAT MEDIATES BINDING TO THE  
 CC POLYMANNOSE O ANTIGEN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

CC EMBL: X69460; G15416; -  
 DR EMBL: AJ001191; E1237676; -  
 DR PIR: S01982; S01982.  
 KW LATE PROTEIN.  
 FT CONFLICT 986 986 V -> A (IN REF. 2).  
 SQ SEQUENCE 1396 AA; 147989 MW; 78DBBAC CMC32;  
 Query Match 41.9%; Score 65; DB 1; Length 1396;  
 Best Local Similarity 70.0%; Pred. No. 5.32e-01;  
 Matches 7; Conservative 3; Mismatches 0; Indels 0; Gaps 0;  
 DB 922 GQMTTPRNG 931  
 QY 1:|||||:1:1  
 8 GOLLTYFRANG 17  
 RESULT 13  
 ID LECT1MOUSE STANDARD; PRT; 301 AA.  
 AC P24721;  
 DT 01-MAR-1992 (REL. 21, CREATED)  
 DT 01-MAR-1992 (REL. 21, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)  
 DE ASINAGOCYCOPROTEIN RECEPTOR 2 (HEPATIC LECTIN 2) (MHL-2) (ASGPR).  
 GN ASGR2 OR ASGR-2.  
 OS MUS MUSCULUS (MOUSE).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;  
 OC RODENTIA; SCIROGNATHI; MURIDAE; MURINAE; MUS.  
 RN 11  
 RP SEQUENCE FROM N.A.  
 RM MEDLINE: 91027942.  
 RA SANFORD J.P., DOYLE D.;  
 RT Mouse asialoglycoprotein receptor cDNA sequence: conservation of  
 RT receptor genes during mammalian evolution.\*;  
 RL BLOCHIM. BIOPHYS. ACTA 1087:259-261(1990).  
 CC -1- FUNCTION: MEDIATES THE ENDOCYTOSIS OF PLASMA GLYCOPROTEINS TO  
 CC WHICH THE TERMINAL SIALIC ACID RESIDE ON THEIR COMPLEX  
 CC CARBOHYDRATE MOIETIES HAS BEEN REMOVED. THE RECEPTOR RECOGNIZES  
 CC TERMINAL GALACTOSE AND N-ACETYLGLUCOSAMINE UNITS. AFTER LIGAND  
 CC BINDING TO THE RECEPTOR, THE RESULTING COMPLEX IS INTERNALIZED AND  
 CC TRANSPORTED TO A SORTING ORGANELLE, WHERE RECEPTOR AND LIGAND ARE  
 CC DISASSOCIATED. THE RECEPTOR THEN RETURNS TO THE CELL MEMBRANE  
 CC SURFACE.  
 CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN.  
 CC -1- TISSUE SPECIFICITY: EXPRESSED EXCLUSIVELY IN HEPATIC PARENCHYMAL  
 CC CELLS.  
 CC -1- CALCIUM IS REQUIRED FOR LIGAND BINDING.  
 CC -1- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: X53042; G53105; -  
 DR PIR: S13165; S13165.  
 DR MGD: MGI:88082; ASGR2.  
 DR PROSITE: PS00615; C\_TYPE\_LLECTIN\_1; 1.  
 DR PROSITE: PS50041; C\_TYPE\_LLECTIN\_2; 1.  
 DR PFAM: PF00059; lectin\_c; 1.  
 DR HSSP: P06734; 1KIE.  
 KW LECTIN; GLYCOPROTEIN; RECEPTOR; ENDOCYTOSIS; TRANSMEMBRANE;  
 KM CALCINUM; SIGNAL-ANCHOR; PHOSPHORYLATION.  
 FT DOMAIN 1 58 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 59 79 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN).  
 FT DOMAIN 80 301 EXTRACELLULAR (POTENTIAL).  
 FT DOMAIN 169 295 C-TYPE LECTIN (LONG FORM).  
 FT DISULFID 170 181 BY SIMILARITY.  
 FT DISULFID 198 293 BY SIMILARITY.

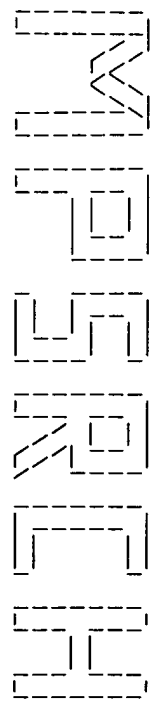
```

CC ROENTIA, SCUROGNATHI, MURIDAE; MURINAE; RATTUS.
RN [1]
RP SEQUENCE FROM N.A.
RA DAVIES J.A., BOCHMAN V., KRYLOVA O., NINKINA N.N.
RT "Molecular cloning and expression pattern of rpr-1, a
RT reseniferatoxin-binding, phosphodiesterase-related protein,
RT expressed in rat kidney tubules.";
RL FEBS LETT. 410:378-382(1997)
CC -1- FUNCTION: BINDS RESINIFEROXIN, A VANILLOID THAT DESENSITISES
CC NOCICEPTIVE NEURONS.
CC -1- COFACTOR: CONTAINS 2 MOLES OF ZINC PER SUBUNIT (BY SIMILARITY).
CC -1- TISSUE SPECIFICITY: EXPRESSED PRIMARILY IN PROXIMAL TUBULES OF THE
CC KIDNEY.
CC -1- SIMILARITY: BELONGS TO THE PHOSPHODIESTERASE FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL; X99477; E254899; -.
CC KM HYDROLASE; ZINC.
CC FT METAL 26 26 ZINC 1 (BY SIMILARITY).
CC FT METAL 28 28 ZINC 1 (BY SIMILARITY).
CC FT METAL 169 169 ZINC 1 AND 2 (BY SIMILARITY).
CC FT METAL 201 201 ZINC 2 (BY SIMILARITY).
CC FT METAL 230 230 ZINC 2 (BY SIMILARITY).
CC FT METAL 298 298 ZINC 1 (BY SIMILARITY).
CC SQ SEQUENCE 349 AA; 38931 MW; A35021FD CRC32;
Query Match 40.6%; Score 63; DB 1; Length 349;
Best Local Similarity 46.2%; Pred. No. 1,27e+00;
Matches 6; Conservative 5; Mismatches 2; Indels 0; Gaps 0;
Db 79 AVRELLYFKAG 91
:::||||:|
QY 5 TINGOLLYFRANG 17

Search completed: Tue Jan 11 15:47:57 2000
Job time : 7 secs

```





Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MSrch\_PP protein - protein database search, using Smith-Waterman algorithm

Run on: Tue Jan 11 15:48:14 2000; MasPar time 5.05 Seconds  
237.695 Million cell updates/sec

Output not generated.

Title: >HS-09-290-049-3

Description: (1-22) from US09290049.pep

Perfect Score: 155

Sequence: 1 TGARTINGQLYFRANGVQVKG 22

Scoring table: PAM 150  
Gap 15

Searched: 179066 seqs, 54579741 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: splrembl9  
1:sp-archaea 2:sp-bacteria 3:sp-fungi 4:sp-human  
5:sp-invertebrate 6:sp-mammal 7:sp-mic 8:sp-organalle  
9:sp-phase 10:sp-plant 11:sp-rodent 12:sp-unclassified  
13:sp-vertebrate 14:sp-virus

Statistics: Mean 29.020; Variance 39.036; scale 0.743  
Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

No.	Score	Match	Length	DB	ID	Description	Pred. No.
1	143	92.3	1476	2	069384	GLUCOSYLTRANSFERASE-I.	5.35e-19
2	143	92.3	1476	2	069396	GLUCOSYLTRANSFERASE-I.	5.35e-19
3	143	92.3	1476	2	069381	GLUCOSYLTRANSFERASE-I.	5.35e-19
4	143	92.3	1476	2	069387	GLUCOSYLTRANSFERASE-I.	5.35e-19
5	143	92.3	1476	2	069390	GLUCOSYLTRANSFERASE-I.	5.35e-19
6	143	92.3	1476	2	069388	GLUCOSYLTRANSFERASE-I.	5.35e-19
7	143	92.3	1476	2	069387	GLUCOSYLTRANSFERASE-I.	5.35e-19
8	143	92.3	1476	2	069387	GLUCOSYLTRANSFERASE-I.	5.35e-19
9	143	92.3	1476	2	069387	GLUCOSYLTRANSFERASE-I.	5.35e-19
10	143	92.3	1476	2	069387	GLUCOSYLTRANSFERASE-I.	5.35e-19
11	143	92.3	1476	2	069387	GLUCOSYLTRANSFERASE-I.	5.35e-19
12	143	92.3	1476	2	069387	GLUCOSYLTRANSFERASE-I.	5.35e-19
13	143	92.3	1476	2	069387	GLUCOSYLTRANSFERASE-I.	5.35e-19
14	143	92.3	1476	2	069387	GLUCOSYLTRANSFERASE-I.	5.35e-19
15	143	92.3	1476	2	069387	GLUCOSYLTRANSFERASE-I.	5.35e-19
16	143	92.3	1476	2	069387	GLUCOSYLTRANSFERASE-I.	5.35e-19
17	143	92.3	1476	2	069387	GLUCOSYLTRANSFERASE-I.	5.35e-19
18	143	92.3	1476	2	069387	GLUCOSYLTRANSFERASE-I.	5.35e-19
19	143	92.3	1476	2	069387	GLUCOSYLTRANSFERASE-I.	5.35e-19
20	143	92.3	1476	2	069387	GLUCOSYLTRANSFERASE-I.	5.35e-19

21	93	60.0	1462	2	069392	GLUCOSYLTRANSFERASE-S.	1.21e-06
22	93	60.0	1462	2	069389	GLUCOSYLTRANSFERASE-S.	1.21e-06
23	93	60.0	1462	2	069389	GLUCOSYLTRANSFERASE-S.	1.21e-06
24	91	58.7	1508	2	052224	GLUCAN-BINDING PROTEIN	1.64e-05
25	88	56.8	1577	2	055265	GLUCAN-BINDING PROTEIN	1.64e-05
26	88	56.8	1577	2	055265	GLUCAN-BINDING PROTEIN	1.64e-05
27	83	53.5	330	2	055228	DEXTRANASE INHIBITOR P	2.08e-04
28	77	49.7	329	1	050120	329A LONG HYPOTHETICA	3.99e-03
29	72	46.5	126	2	047172	DNA SEQUENCE UPSTREAM	4.28e-02
30	72	46.5	126	2	047172	ANALOGUE.	4.28e-02
31	72	46.5	126	2	047171	ANALOGUE.	4.28e-02
32	71	45.8	425	2	067315	UDP-N-ACETYLGLUCOSAMIN	6.80e-02
33	68	43.9	831	2	050076	CHITINASE B (EC 3.2.1.	2.67e-01
34	67	43.2	245	1	058723	HYPOTHETICAL PROTEIN M	4.18e-01
35	65	41.9	446	11	063237	HEPARIN-BINDING FIBROB	1.01e+00
36	65	41.9	446	11	063236	HEPARIN-BINDING FIBROB	1.01e+00
37	64	41.3	221	5	044353	GATA TRANSCRIPTION FAC	1.56e+00
38	64	41.3	669	10	081752	HYPOTHETICAL 75.2 KD P	1.56e+00
39	64	41.3	829	5	023290	SIMILAR TO SERINE/THRE	1.56e+00
40	63	40.6	240	1	029326	CONSERVED HYPOTHETICAL	2.40e+00
41	63	40.6	371	10	081753	HYPOTHETICAL 42.3 KD P	2.40e+00
42	63	40.6	566	2	P72770	HIGH AFFINITY SULFATE	2.40e+00
43	63	40.6	574	10	081754	HYPOTHETICAL 65.1 KD P	2.40e+00
44	62	40.0	1360	5	094649	AARF2 PROTEIN.	3.68e+00
45	62	40.0	2364	2	046342	CYTOTOXIN L.	3.68e+00

ALIGNMENTS

RESULT	ID	PRELIMINARY	PRT	1476 AA.
1	069384			
AC	069384			
DT	01-AUG-1998 (TREMBLREL. 07, CREATED)			
DT	01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)			
DT	01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)			
DE	GLUCOSYLTRANSFERASE-I.			
GN	GTFB.			
OS	STREPTOCOCCUS MUTANS.			
OC	BACILLUS/CLOSTRIDIUM GROUP; STREPTOCOCCACEAE;			
OC	STREPTOCOCCUS.			
RN	(1)			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-MT4245.			
RC	MEDLINE: 98231643.			
RA	FUJIMURA T., TERAO Y., HOSHINO T., KAMABATA S., OOSHIMA T., SOBUE S.,			
RA	KIMURA S., HAMADA S.,			
RT	*Molecular analyses of glucosyltransferase genes among strains of			
RT	streptococcus mutans. "			
RT	FEWS MICROBIOL. LETT. 161:331-336(1998).			
RT	EMBL: D88654; D1027045; -			
KW	TRANSFERASE.			
SO	SEQUENCE 1476 AA; 165792 MW; C76D8483 CRC32;			

Query Match 92.3%; Score 143; DB 2; Length 1476;  
Best Local Similarity 95.5%; Pred. No. 5.35e-19;  
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

DB	1301	TGARTINGQLYFRANGVQVKG 1322
QY	1	TGARTINGQLYFRANGVQVKG 22
RESULT	2	
ID	069396	PRELIMINARY: PRT: 1476 AA.
AC	069396	
DT	01-AUG-1998 (TREMBLREL. 07, CREATED)	
DT	01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)	
DT	01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)	
DE	GLUCOSYLTRANSFERASE-I.	
GN	GTFB.	
OS	STREPTOCOCCUS MUTANS.	
OC	BACILLUS/CLOSTRIDIUM GROUP; STREPTOCOCCACEAE;	
OC	STREPTOCOCCUS.	

```

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-MT4467;
RX MEDLINE: 98231643.
RA FUJIMURA T., TERAO Y., HOSHINO T., KAWABATA S., OOSHIMA T., SOBUE S.,
RA KIMURA S., HAMADA S.;
RT "Molecular analyses of glucosyltransferase genes among strains of
RT Streptococcus mutans.";
RL FEMS MICROBIOL. LETT. 161:331-336(1998).
DR EMBL: D89977; D1027059; -.
KW TRANSFERASE.
SO SEQUENCE 1476 AA; 165647 MW; 258D4ADD CRC32;

Query Match          92.3%; Score 143; DB 2; Length 1476;
Best Local Similarity 95.5%; Pred. No. 5.35e-19;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 1301 TGARTINGOLLYFRANGOVYKG 1322
1 TGARTINGOLLYFRANGOVYKG 22
|||||

RESULT 3
ID 069381; PRELIMINARY; PRT; 1476 AA.
AC 069381;
DT 01-AUG-1998 (TREMBLREL. 07, CREATED)
DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE GLUCOSYLTRANSFERASE-I.
GN GTFB.
OS STREPTOCOCCUS MUTANS.
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; STREPTOCOCCACEAE;
OC STREPTOCOCCUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-MT8148;
RX MEDLINE: 98231643.
RA FUJIMURA T., TERAO Y., HOSHINO T., KAWABATA S., OOSHIMA T., SOBUE S.,
RA KIMURA S., HAMADA S.;
RT "Molecular analyses of glucosyltransferase genes among strains of
RT Streptococcus mutans.";
RL FEMS MICROBIOL. LETT. 161:331-336(1998).
DR EMBL: D88651; D1027041; -.
KW TRANSFERASE.
SO SEQUENCE 1476 AA; 165638 MW; 0868C20A CRC32;

Query Match          92.3%; Score 143; DB 2; Length 1476;
Best Local Similarity 95.5%; Pred. No. 5.35e-19;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 1301 TGARTINGOLLYFRANGOVYKG 1322
1 TGARTINGOLLYFRANGOVYKG 22
|||||

RESULT 4
ID 069387; PRELIMINARY; PRT; 1476 AA.
AC 069387;
DT 01-AUG-1998 (TREMBLREL. 07, CREATED)
DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE GLUCOSYLTRANSFERASE-I.
GN GTFB.
OS STREPTOCOCCUS MUTANS.
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; STREPTOCOCCACEAE;
OC STREPTOCOCCUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-MT4239;
RX MEDLINE: 98231643.
RA FUJIMURA T., TERAO Y., HOSHINO T., KAWABATA S., OOSHIMA T., SOBUE S.,
RA KIMURA S., HAMADA S.;
RT "Molecular analyses of glucosyltransferase genes among strains of
RT Streptococcus mutans.";
RL FEMS MICROBIOL. LETT. 161:331-336(1998).
DR EMBL: D88658; D1027050; -.
KW TRANSFERASE.
SO SEQUENCE 1455 AA; 163046 MW; 7659CFBC CRC32;

Query Match          92.3%; Score 143; DB 2; Length 1476;
Best Local Similarity 95.5%; Pred. No. 5.35e-19;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 1301 TGARTINGOLLYFRANGOVYKG 1322
1 TGARTINGOLLYFRANGOVYKG 22
|||||

RESULT 5
ID 069390; PRELIMINARY; PRT; 1476 AA.
AC 069390;
DT 01-AUG-1998 (TREMBLREL. 07, CREATED)
DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE GLUCOSYLTRANSFERASE-I.
GN GTFB.
OS STREPTOCOCCUS MUTANS.
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; STREPTOCOCCACEAE;
OC STREPTOCOCCUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-MT4251;
RX MEDLINE: 98231643.
RA FUJIMURA T., TERAO Y., HOSHINO T., KAWABATA S., OOSHIMA T., SOBUE S.,
RA KIMURA S., HAMADA S.;
RT "Molecular analyses of glucosyltransferase genes among strains of
RT Streptococcus mutans.";
RL FEMS MICROBIOL. LETT. 161:331-336(1998).
DR EMBL: D88660; D1027053; -.
KW TRANSFERASE.
SO SEQUENCE 1476 AA; 165833 MW; 9119C72B CRC32;

Query Match          92.3%; Score 143; DB 2; Length 1476;
Best Local Similarity 95.5%; Pred. No. 5.35e-19;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 1301 TGARTINGOLLYFRANGOVYKG 1322
1 TGARTINGOLLYFRANGOVYKG 22
|||||

RESULT 6
ID 069388; PRELIMINARY; PRT; 1455 AA.
AC 069388;
DT 01-AUG-1998 (TREMBLREL. 07, CREATED)
DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE GLUCOSYLTRANSFERASE-SI.
GN GTFB.
OS STREPTOCOCCUS MUTANS.
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; STREPTOCOCCACEAE;
OC STREPTOCOCCUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-MT4239;
RX MEDLINE: 98231643.
RA FUJIMURA T., TERAO Y., HOSHINO T., KAWABATA S., OOSHIMA T., SOBUE S.,
RA KIMURA S., HAMADA S.;
RT "Molecular analyses of glucosyltransferase genes among strains of
RT Streptococcus mutans.";
RL FEMS MICROBIOL. LETT. 161:331-336(1998).
DR EMBL: D88658; D1027050; -.
KW TRANSFERASE.
SO SEQUENCE 1455 AA; 163046 MW; 7659CFBC CRC32;

Query Match          78.1%; Score 121; DB 2; Length 1455;
```

```

RT Streptococcus mutans.";
RL FEMS MICROBIOL. LETT. 161:331-336(1998).
DR EMBL: D88657; D1027049; -.
KW TRANSFERASE.
SO SEQUENCE 1476 AA; 165819 MW; AF6DBA6F CRC32;

Query Match          92.3%; Score 143; DB 2; Length 1476;
Best Local Similarity 95.5%; Pred. No. 5.35e-19;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 1301 TGARTINGOLLYFRANGOVYKG 1322
1 TGARTINGOLLYFRANGOVYKG 22
|||||

RESULT 5
ID 069390; PRELIMINARY; PRT; 1476 AA.
AC 069390;
DT 01-AUG-1998 (TREMBLREL. 07, CREATED)
DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE GLUCOSYLTRANSFERASE-I.
GN GTFB.
OS STREPTOCOCCUS MUTANS.
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; STREPTOCOCCACEAE;
OC STREPTOCOCCUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-MT4251;
RX MEDLINE: 98231643.
RA FUJIMURA T., TERAO Y., HOSHINO T., KAWABATA S., OOSHIMA T., SOBUE S.,
RA KIMURA S., HAMADA S.;
RT "Molecular analyses of glucosyltransferase genes among strains of
RT Streptococcus mutans.";
RL FEMS MICROBIOL. LETT. 161:331-336(1998).
DR EMBL: D88660; D1027053; -.
KW TRANSFERASE.
SO SEQUENCE 1476 AA; 165833 MW; 9119C72B CRC32;

Query Match          92.3%; Score 143; DB 2; Length 1476;
Best Local Similarity 95.5%; Pred. No. 5.35e-19;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 1301 TGARTINGOLLYFRANGOVYKG 1322
1 TGARTINGOLLYFRANGOVYKG 22
|||||

RESULT 6
ID 069388; PRELIMINARY; PRT; 1455 AA.
AC 069388;
DT 01-AUG-1998 (TREMBLREL. 07, CREATED)
DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE GLUCOSYLTRANSFERASE-SI.
GN GTFB.
OS STREPTOCOCCUS MUTANS.
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; STREPTOCOCCACEAE;
OC STREPTOCOCCUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-MT4239;
RX MEDLINE: 98231643.
RA FUJIMURA T., TERAO Y., HOSHINO T., KAWABATA S., OOSHIMA T., SOBUE S.,
RA KIMURA S., HAMADA S.;
RT "Molecular analyses of glucosyltransferase genes among strains of
RT Streptococcus mutans.";
RL FEMS MICROBIOL. LETT. 161:331-336(1998).
DR EMBL: D88658; D1027050; -.
KW TRANSFERASE.
SO SEQUENCE 1455 AA; 163046 MW; 7659CFBC CRC32;

Query Match          78.1%; Score 121; DB 2; Length 1455;
```





ID 000600 PRELIMINARY; PRT; 1518 AA.  
 AC 000600;  
 DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
 DE GLUCOSYLTRANSFERASE I (EC 2.4.1.5) (GTF) (DEXTRANSUCRASE).  
 DE (SUCROSE 6-GLUCOSYLTRANSFERASE).  
 GN GTF.  
 OS STREPTOCOCCUS SALIVARIUS.  
 OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; STREPTOCOCCACEAE;  
 OC STREPTOCOCCUS.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-ATCC 25975;  
 RX MEDLINE: 92148377.  
 RA GIFFARD P.M., SIMPSON C.L., MILWARD C.P., JACQUES N.A.;  
 RT "Molecular characterization of a cluster of at least two  
 glucosyltransferase genes in Streptococcus salivarius ATCC 25975.";  
 J. GEN. MICROBIOL. 137:2577-2593(1991).  
 CC -1- FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS, THAT ARE THOUGHT  
 TO PLAY A KEY ROLE IN THE DEVELOPMENT OF THE DENTAL PLAQUE  
 BECAUSE OF THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDIATE  
 THE AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS.  
 CC -1- CATALYTIC ACTIVITY: SUCROSE + D-FRUCTOSE +  
 (1,6-ALPHA-D-GLUCOSYL)(N) - D-FRUCTOSE +  
 (1,6-ALPHA-D-GLUCOSYL)(N+1).  
 CC -1- SUBCELLULAR LOCATION: EXTRACELLULAR.  
 CC -1- DISEASE: DENTAL CARIES.  
 DR EMBL: Z11873; G47527; -;  
 DR EMBL: M64111; G153649; -;  
 DR PRAM: PF00128; alpha-amylase; 1.  
 KW TRANSFERASE; GLUCOSYLTRANSFERASE; REPEAT; DENTAL CARIES.  
 FT DOMAIN 1307 1482 6 DIRECT REPEATS.  
 FT REPEAT 1307 1338 REPEAT 1.  
 FT REPEAT 1339 1352 REPEAT 2.  
 FT REPEAT 1372 1403 REPEAT 3.  
 FT REPEAT 1404 1417 REPEAT 4.  
 FT REPEAT 1437 1468 REPEAT 5.  
 FT REPEAT 1469 1482 REPEAT 6.  
 SQ SEQUENCE 1518 AA; 167730 MW; 20728800 CRC32;

Query Match 70.3%; Score 109; DB 2; Length 1518;  
 Best Local Similarity 72.7%; Pred. No. 2.05e-10;  
 Matches 16; Conservative 2; Mismatches 4; Indels 0; Gaps 0;  
 Db 1383 TGAOVINGOHLFYENADGSOVKG 1404  
 |||:||||| |||:|:|||||  
 1 TGARTINGOLLYFRANGVOYKG 22

RESULT 12  
 ID 055263 PRELIMINARY; PRT; 1590 AA.  
 AC 055263;  
 DT 01-JAN-1996 (TREMBLREL. 01, CREATED)  
 DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
 DE GLUCOSYLTRANSFERASE GTF-1.  
 OS STREPTOCOCCUS SOBRINUS.  
 OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; STREPTOCOCCACEAE;  
 OC STREPTOCOCCUS.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-ATCC 33478;  
 RA SATO S.;  
 RA ANN. KAGOSHIMA UNIV. DENTAL SCHOOL 16:23-29(1996).  
 DR EMBL: D63570; D1010438; -;  
 DR PRAM: PF00128; alpha-amylase; 1.  
 KW TRANSFERASE.  
 SQ SEQUENCE 1590 AA; 176057 MW; 471866A CRC32;

Query Match 70.3%; Score 109; DB 2; Length 1590;  
 Best Local Similarity 68.2%; Pred. No. 2.05e-10;  
 Matches 15; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Db 1410 TGEOTINGOHLFYKEDGHQYKG 1431  
 ||:||||| |||:|:|||||  
 1 TGARTINGOLLYFRANGVOYKG 22

RESULT 13  
 ID 048756 PRELIMINARY; PRT; 1290 AA.  
 AC 048756;  
 DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
 DE DEXTRANSUCRASE.  
 OS LEUCONOSTOC MESENEROIDES.  
 OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; LACTOBACILLACEAE;  
 OC LEUCONOSTOC.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-NRRL B1299;  
 RX MEDLINE: 97136686.  
 RA MONCHOIS V., WILLEMOT R.M., REMAUD-SIMEON M., CROUX C., MONSAN P.;  
 RT "Cloning and sequencing of a gene coding for a novel dextranucrase  
 from Leuconostoc mesenteroides NRRL B-1299 synthesizing only alpha  
 (1-6) and alpha (1-3) linkages.";  
 RL GENE 182:23-32(1996).  
 DR EMBL: U38181; G1022963; -;  
 DR PRAM: PF00128; alpha-amylase; 1.  
 SQ SEQUENCE 1290 AA; 145590 MW; F4460B13 CRC32;

Query Match 69.0%; Score 107; DB 2; Length 1290;  
 Best Local Similarity 63.6%; Pred. No. 6.24e-10;  
 Matches 14; Conservative 3; Mismatches 5; Indels 0; Gaps 0;  
 Db 1181 TGDRIINGOKLYPNONGIOMKG 1202  
 ||:||||| |||:|:|||||  
 1 TGARTINGOLLYFRANGVOYKG 22

RESULT 14  
 ID 069385 PRELIMINARY; PRT; 1390 AA.  
 AC 069385;  
 DT 01-APR-1998 (TREMBLREL. 07, CREATED)  
 DT 01-APR-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
 DE GLUCOSYLTRANSFERASE-SI.  
 GN GTF.  
 OS STREPTOCOCCUS MUTANS.  
 OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; STREPTOCOCCACEAE;  
 OC STREPTOCOCCUS.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-MT4245;  
 RX MEDLINE: 98231643.  
 RA FUJIMURA T., TERAO Y., HOSHINO T., KAWABATA S., OOSHIMA T., SOBE S.,  
 RA KIMURA S., HAMADA S.;  
 RT "Molecular analyses of glucosyltransferase genes among strains of  
 Streptococcus mutans.";  
 RT FEMS MICROBIOL. LETT. 161:331-336(1998).  
 DR EMBL: D88655; D1027046; -;  
 KW TRANSFERASE.  
 SQ SEQUENCE 1390 AA; 155375 MW; C211E7B8 CRC32;

Query Match 68.4%; Score 106; DB 2; Length 1390;  
 Best Local Similarity 68.2%; Pred. No. 1.09e-09;  
 Matches 15; Conservative 4; Mismatches 3; Indels 0; Gaps 0;  
 Db 1264 TGTVTENGORLYFKPNQOAKG 1285  
 ||:||||| |||:|:|||||  
 1 TGARTINGOLLYFRANGVOYKG 22

RESULT 15  
 ID 000599 PRELIMINARY; PRT; 1599 AA.

AC 000599;  
 DT 01-NOV-1996 (TREMELREL. 01, CREATED)  
 DT 01-NOV-1996 (TREMELREL. 01, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1998 (TREMELREL. 08, LAST ANNOTATION UPDATE)  
 DE GLUCOSYLTRANSFERASE S PRECURSOR (EC 2.4.1.5) (GTF) (DEXTRANSUCRASE)  
 DE (SUCROSE 6-GLUCOSYLTRANSFERASE).  
 GN GTF.  
 OS STREPTOCOCCUS SALIVARIUS.  
 OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; STREPTOCOCCACEAE;  
 OC STREPTOCOCCUS.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-ATCC 25975;  
 RX MEDLINE; 9381463.  
 RA GIFFARD P.M., ALLEN D.M., MILWARD C.P., SIMPSON C.L., JACQUES N.A.;  
 RT "Sequence of the gtf gene of Streptococcus salivarius ATCC 25975 and  
 evolution of the gtf genes of oral streptococci.";  
 RL J. GEN. MICROBIOL. 139:1511-1522(1993).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-ATCC 25975;  
 RX MEDLINE; 92148377.  
 RA GIFFARD P.M., SIMPSON C.L., MILWARD C.P., JACQUES N.A.;  
 RT "Molecular characterization of a cluster of at least two  
 glucosyltransferase genes in Streptococcus salivarius ATCC 25975.";  
 RL J. GEN. MICROBIOL. 137:2577-2593(1991).  
 CC -1- FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS, THAT ARE THOUGHT  
 TO PLAY A KEY ROLE IN THE DEVELOPMENT OF THE DENTAL PLAQUE  
 BECAUSE OF THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDIATE  
 THE AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS.  
 CC -1- CATALYTIC ACTIVITY: SUCROSE +  
 (1,6-ALPHA-D-GLUCOSYL) (N) - D-FRUCTOSE +  
 (1,6-ALPHA-D-GLUCOSYL) (N+1).  
 CC -1- SUBCELLULAR LOCATION: EXTRACELLULAR.  
 CC -1- DISEASE: DENTAL CARIES.  
 CC -1- SIMILARITY: TO REGIONS OF BARLEY AND BACILLUS AMYLOLIQUEFACIENS  
 CC ALPHA AMYLASES AND RABBIT GLYCOGEN PHOSPHORYLASE.  
 DR EMBL; 211872; G47531; -;  
 DR EMBL; 211873; G47528; -;  
 DR EMBL; M6411; G153650; -;  
 DR PIRAM; PF00128; alpha-amylase; 1.  
 KW TRANSFERASE; GLYCOSYLTRANSFERASE; SIGNAL; REPEAT; DENTAL CARIES.  
 FT SIGNAL 1 42  
 FT CHAIN 43 1599  
 FT SIMILAR 867 967  
 FT SIMILAR 1132 1219  
 SQ SEQUENCE 1599 AA; 176480 MW; 575C79C9 CRC32;  
 Query Match 68.4%; Score 106; DB 2; Length 1599;  
 st Local Similarity 77.3%; Pred. No. 1.09e-09;  
 tches 17; Conservative 1; Mismatches 4; Indels 0; Gaps 0;  
 DB 1402 TGAOVINGHLEFDANGROYKG 1423  
 OY 1 TGAFTINGQLTYFRANGVOYKG 22

Search completed: Tue Jan 11 15:49:06 2000  
 Job time : 52 secs.



\*\*\*\*\*  
 MIPSETH  
 (TM)  
 \*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
 Copyright (c) 1993-1998 University of Edinburgh, U.K.  
 Distribution rights by Oxford Molecular Ltd

Mpsrch\_dp protein - protein database search, using Smith-Waterman algorithm

Run on: Tue Jan 11 15:37:52 2000; Maspar time 5.64 Seconds 203.291 Million cell updates/sec

\*\*\*\*\*  
 T: ar output not generated.

Title: >US-09-290-049-1

Description: (1-21) from US09290049.pep

Perfect Score: 148

Sequence: 1 ANDHLSILEAWSNDNTPYLHD 21

Scoring table: PAM 150  
 Gap 15

Searched: 179066 segs, 54579741 residues

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database:

sptrembl9  
 1:sp\_archaea 2:sp\_bacteria 3:sp\_fungi 4:sp\_human  
 5:sp\_invertebrate 6:sp\_mammal 7:sp\_mhc 8:sp\_organelle  
 9:sp\_phase 10:sp\_plant 11:sp\_rodent 12:sp\_unclassified  
 13:sp\_vertebrate 14:sp\_virus

Statistics: Mean 28.660; Variance 41.580; scale 0.689

Pred. No. is the number of results predicted by chance to have a  
 score greater than or equal to the score of the result being printed,  
 and is derived by analysis of the total score distribution.

## SUMMARIES

Hit No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	148	100.0	1390	2	069385	GLUCOSYLTTRANSFERASE-ST
2	148	100.0	1455	2	069391	GLUCOSYLTTRANSFERASE-ST
3	148	100.0	1476	2	069390	GLUCOSYLTTRANSFERASE-ST
4	148	100.0	1476	2	069381	GLUCOSYLTTRANSFERASE-ST
5	148	100.0	1476	2	069384	GLUCOSYLTTRANSFERASE-ST
6	148	100.0	1476	2	069396	GLUCOSYLTTRANSFERASE-ST
7	148	100.0	1476	2	069387	GLUCOSYLTTRANSFERASE-ST
8	136	91.9	1455	2	069397	GLUCOSYLTTRANSFERASE-ST
9	136	91.9	1455	2	069388	GLUCOSYLTTRANSFERASE-ST
10	136	91.9	1455	2	069382	GLUCOSYLTTRANSFERASE-ST
11	131	88.5	1590	2	055263	GLUCOSYLTTRANSFERASE-ST
12	131	88.5	1590	2	059983	GLUCOSYLTTRANSFERASE-ST
13	101	68.2	1462	2	069398	GLUCOSYLTTRANSFERASE-ST
14	101	68.2	1462	2	069386	GLUCOSYLTTRANSFERASE-ST
15	101	68.2	1462	2	069389	GLUCOSYLTTRANSFERASE-ST
16	101	68.2	1462	2	069392	GLUCOSYLTTRANSFERASE-ST
17	101	68.2	1462	2	069383	GLUCOSYLTTRANSFERASE-ST
18	101	68.2	1508	2	052224	GLUCOSYLTTRANSFERASE-ST
19	95	64.2	1577	2	054178	GLUCOSYLTTRANSFERASE-ST
20	86	58.1	1449	2	055264	GLUCOSYLTTRANSFERASE-ST

21	86	58.1	1449	2	068542	GLUCOSYLTTRANSFERASE N	1.41e-04
22	83	56.1	1577	2	055265	GLUCOSYLTTRANSFERASE PR	5.84e-04
23	78	52.7	1599	2	000599	GLUCOSYLTTRANSFERASE S	5.88e-03
24	75	50.7	1518	2	000600	GLUCOSYLTTRANSFERASE I	2.27e-02
25	72	48.6	579	2	084087	HYPOHETICAL 65.4 KD P	8.55e-02
26	70	47.3	549	10	043274	T CYTOPLASM MALE STER	2.03e-01
27	70	47.3	601	2	005566	PPDK.	2.03e-01
28	68	45.9	175	14	P89063	NSP4.	4.77e-01
29	68	45.9	175	14	P89062	NSP4.	4.77e-01
30	67	45.3	490	2	006579	HYPOHETICAL 52.0 KD P	7.26e-01
31	67	45.3	525	3	059679	HYPOHETICAL 60.3 KD P	7.26e-01
32	66	44.6	464	10	004073	PHOTOSYSTEM II DI PROT	1.10e+00
33	66	44.6	4848	2	007944	PRISTINAMYCIN I SYNTHA	1.10e+00
34	65	43.9	581	2	034023	HYPOHETICAL 66.2 KD P	1.66e+00
35	64	43.2	515	10	023614	PSII D1 PROTEIN PROCES	2.50e+00
36	64	43.2	542	10	P93344	ALDEHYDE DEHYDROGENASE	2.50e+00
37	64	43.2	594	14	091465	CHITINASE.	2.50e+00
38	63	42.6	175	14	056850	NSP4.	3.75e+00
39	63	42.6	175	14	056851	NSP4.	3.75e+00
40	63	42.6	403	10	023222	MADS-BOX PROTEIN HOMOL	3.75e+00
41	62	41.9	360	5	045983	ZC513.10 PROTEIN.	5.59e+00
42	62	41.9	380	14	085633	V-MIL PROTEIN.	5.59e+00
43	62	41.9	420	13	090893	C-MIL PHOTO-ONCOGENE (	5.59e+00
44	62	41.9	602	11	063485	RAT ACTIVATED C-RAF ON	5.59e+00
45	62	41.9	1009	5	017380	F5263.4 PROTEIN.	5.59e+00

## ALIGNMENTS

RESULT	ID	PRELIMINARY	PRT	1390 AA.
AC	069385			
AD	069385			
DI	01-AUG-1998 (TREMBLREL. 07, CREATED)			
DI	01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)			
DI	01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)			
DE	GLUCOSYLTTRANSFERASE-ST.			
GN	GTEC.			
OS	STREPTOCOCCUS MUTANS.			
OC	BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; STREPTOCOCCACEAE;			
OC	STREPTOCOCCUS.			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-MT4245.			
RC	MEDLINE; 98231643.			
RA	FUJIMURA T., TERAO Y., HOSHINO T., KAWABATA S., OOSHIMA T., SOBUE S.,			
RA	KIMURA S., HAMADA S.,			
RT	"Molecular analyses of glucosyltransferase genes among strains of			
RT	Streptococcus mutans."			
RT	FEMS MICROBIOL. LETT. 161:331-336(1998).			
DR	EMBL; D88655; D1027046; -			
KM	TRANSFERASE.			
SO	SEQUENCE 1390 AA; 155375 MW; C211E7B8 CRC32;			

Query Match 100.0%; Score 148; DB 2; Length 1390;

Best Local Similarity 100.0%; Pred. No. 8.69e-19;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 507 ANDHLSILEAWSNDNTPYLHD 527  
 QY 1 ANDHLSILEAWSNDNTPYLHD 21

RESULT 2 PRELIMINARY: PRT: 1455 AA.

AC 069391.  
 DI 01-AUG-1998 (TREMBLREL. 07, CREATED)  
 DI 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)  
 DI 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
 DE GLUCOSYLTTRANSFERASE-ST.  
 GN GTEC.  
 OS STREPTOCOCCUS MUTANS.  
 OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; STREPTOCOCCACEAE;  
 OC STREPTOCOCCUS.

RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-MT4251;  
 RX MEDLINE: 98231643.  
 RA FUJIMURA T., TERAO Y., HOSHINO T., KAWABATA S., OOSHIMA T., SOBUE S.,  
 RA KIMURA S., HANADA S.;  
 RT "Molecular analyses of glucosyltransferase genes among strains of  
 RT Streptococcus mutans.";  
 RL FEMS MICROBIOL. LETT. 161:331-336(1998).  
 DR EMBL: D88651; D1027054; -  
 KW TRANSFERASE.  
 SQ SEQUENCE 1455 AA; 162804 MW; DB730514 CRC32;

Query Match 100.0%; Score 148; DB 2; Length 1455;  
 Best Local Similarity 100.0%; Pred. No. 8.69e-19;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 507 ANDHLSILEMSDNDTPYLHD 527  
 1 ANDHLSILEMSDNDTPYLHD 21

RESULT 3  
 ID 069390 PRELIMINARY; PRT: 1476 AA.  
 AC 069390;  
 DT 01-AUG-1998 (TREMBLREL. 07, CREATED)  
 DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
 DE GLUCOSYLTRANSFERASE-I.  
 GN GTFB.  
 OS STREPTOCOCCUS MUTANS.  
 OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; STREPTOCOCCACEAE;  
 OC STREPTOCOCCUS.  
 [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-MT4251;  
 RX MEDLINE: 98231643.  
 RA FUJIMURA T., TERAO Y., HOSHINO T., KAWABATA S., OOSHIMA T., SOBUE S.,  
 RA KIMURA S., HANADA S.;  
 RT "Molecular analyses of glucosyltransferase genes among strains of  
 RT Streptococcus mutans.";  
 RL FEMS MICROBIOL. LETT. 161:331-336(1998).  
 DR EMBL: D88650; D1027053; -  
 KW TRANSFERASE.  
 SQ SEQUENCE 1476 AA; 165833 MW; 9119C72B CRC32;

Query Match 100.0%; Score 148; DB 2; Length 1476;  
 Best Local Similarity 100.0%; Pred. No. 8.69e-19;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 481 ANDHLSILEMSDNDTPYLHD 501  
 1 ANDHLSILEMSDNDTPYLHD 21

RESULT 4  
 ID 069381 PRELIMINARY; PRT: 1476 AA.  
 AC 069381;  
 DT 01-AUG-1998 (TREMBLREL. 07, CREATED)  
 DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
 DE GLUCOSYLTRANSFERASE-I.  
 GN GTFB.  
 OS STREPTOCOCCUS MUTANS.  
 OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; STREPTOCOCCACEAE;  
 OC STREPTOCOCCUS.  
 [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-MT8148;  
 RX MEDLINE: 98231643.  
 RA FUJIMURA T., TERAO Y., HOSHINO T., KAWABATA S., OOSHIMA T., SOBUE S.,  
 RA KIMURA S., HANADA S.;  
 RT "Molecular analyses of glucosyltransferase genes among strains of

RT Streptococcus mutans.";  
 RL FEMS MICROBIOL. LETT. 161:331-336(1998).  
 DR EMBL: D88651; D1027041; -  
 KW TRANSFERASE.  
 SQ SEQUENCE 1476 AA; 165638 MW; 0868C20A CRC32;

Query Match 100.0%; Score 148; DB 2; Length 1476;  
 Best Local Similarity 100.0%; Pred. No. 8.69e-19;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 481 ANDHLSILEMSDNDTPYLHD 501  
 1 ANDHLSILEMSDNDTPYLHD 21

RESULT 5  
 ID 069384 PRELIMINARY; PRT: 1476 AA.  
 AC 069384;  
 DT 01-AUG-1998 (TREMBLREL. 07, CREATED)  
 DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
 DE GLUCOSYLTRANSFERASE-I.  
 GN GTFB.  
 OS STREPTOCOCCUS MUTANS.  
 OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; STREPTOCOCCACEAE;  
 OC STREPTOCOCCUS.  
 [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-MT4245;  
 RX MEDLINE: 98231643.  
 RA FUJIMURA T., TERAO Y., HOSHINO T., KAWABATA S., OOSHIMA T., SOBUE S.,  
 RA KIMURA S., HANADA S.;  
 RT "Molecular analyses of glucosyltransferase genes among strains of  
 RT Streptococcus mutans.";  
 RL FEMS MICROBIOL. LETT. 161:331-336(1998).  
 DR EMBL: D88654; D1027045; -  
 KW TRANSFERASE.  
 SQ SEQUENCE 1476 AA; 165792 MW; C76D8483 CRC32;

Query Match 100.0%; Score 148; DB 2; Length 1476;  
 Best Local Similarity 100.0%; Pred. No. 8.69e-19;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 481 ANDHLSILEMSDNDTPYLHD 501  
 1 ANDHLSILEMSDNDTPYLHD 21

RESULT 6  
 ID 069396 PRELIMINARY; PRT: 1476 AA.  
 AC 069396;  
 DT 01-AUG-1998 (TREMBLREL. 07, CREATED)  
 DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
 DE GLUCOSYLTRANSFERASE-I.  
 GN GTFB.  
 OS STREPTOCOCCUS MUTANS.  
 OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; STREPTOCOCCACEAE;  
 OC STREPTOCOCCUS.  
 [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-MT4467;  
 RX MEDLINE: 98231643.  
 RA FUJIMURA T., TERAO Y., HOSHINO T., KAWABATA S., OOSHIMA T., SOBUE S.,  
 RA KIMURA S., HANADA S.;  
 RT "Molecular analyses of glucosyltransferase genes among strains of  
 RT Streptococcus mutans.";  
 RL FEMS MICROBIOL. LETT. 161:331-336(1998).  
 DR EMBL: D89977; D1027059; -  
 KW TRANSFERASE.  
 SQ SEQUENCE 1476 AA; 165647 MW; 258D4ADD CRC32;

Query Match 100.0%; Score 148; DB 2; Length 1476;

Best Local Similarity 100.0%; Pred. No. 8.69e-19;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 481 ANDHLSILEAMSDNDTPYLHD 501  
|||||  
QY 1 ANDHLSILEAMSDNDTPYLHD 21

RESULT 7  
ID 069387 PRELIMINARY; PRT; 1476 AA.  
AC 069387;  
DT 01-AUG-1998 (TREMBLREL. 07, CREATED)  
DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)  
DE 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
DE GLUCOSYLTRANSFERASE-I.  
GN GTFB.  
OS STREPTOCOCCUS MUTANS.  
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; STREPTOCOCCACEAE;  
OC STREPTOCOCCUS.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN-MT4239;  
RX MEDLINE: 98231643.  
RA FUJIWARA T., TERAO Y., HOSHINO T., KAWABATA S., OOSHIMA T., SOBUE S.,  
RA KIMURA S., HAMADA S.;  
RT "Molecular analyses of glucosyltransferase genes among strains of  
RT streptococcus mutans.";  
RL FEMS MICROBIOL. LETT. 161:331-336(1998).  
DR EMBL: D88657; D1027049; -.  
KW TRANSFERASE.  
SQ SEQUENCE 1476 AA; 165819 MW; AF6DBA6F CRC32;

Query Match 100.0%; Score 148; DB 2; Length 1476;  
Best Local Similarity 100.0%; Pred. No. 8.69e-19;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 481 ANDHLSILEAMSDNDTPYLHD 501  
|||||  
QY 1 ANDHLSILEAMSDNDTPYLHD 21

RESULT 8  
ID 069397 PRELIMINARY; PRT; 1455 AA.  
AC 069397;  
DT 01-AUG-1998 (TREMBLREL. 07, CREATED)  
DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)  
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
DE GLUCOSYLTRANSFERASE-SI.  
GN GTFB.  
OS STREPTOCOCCUS MUTANS.  
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; STREPTOCOCCACEAE;  
OC STREPTOCOCCUS.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN-MT4467;  
RX MEDLINE: 98231643.  
RA FUJIWARA T., TERAO Y., HOSHINO T., KAWABATA S., OOSHIMA T., SOBUE S.,  
RA KIMURA S., HAMADA S.;  
RT "Molecular analyses of glucosyltransferase genes among strains of  
RT streptococcus mutans.";  
RL FEMS MICROBIOL. LETT. 161:331-336(1998).  
DR EMBL: D89978; D1027060; -.  
KW TRANSFERASE.  
SQ SEQUENCE 1455 AA; 162914 MW; 7D729DA0 CRC32;

Query Match 91.9%; Score 136; DB 2; Length 1455;  
Best Local Similarity 95.2%; Pred. No. 7.09e-16;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

DB 507 ANDHLSILEAMSDNDTPYLHD 527  
|||||  
QY 1 ANDHLSILEAMSDNDTPYLHD 21

RESULT 9  
ID 069388 PRELIMINARY; PRT; 1455 AA.  
AC 069388;  
DT 01-AUG-1998 (TREMBLREL. 07, CREATED)  
DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)  
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
DE GLUCOSYLTRANSFERASE-SI.  
GN GTFB.  
OS STREPTOCOCCUS MUTANS.  
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; STREPTOCOCCACEAE;  
OC STREPTOCOCCUS.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN-MT4239;  
RX MEDLINE: 98231643.  
RA FUJIWARA T., TERAO Y., HOSHINO T., KAWABATA S., OOSHIMA T., SOBUE S.,  
RA KIMURA S., HAMADA S.;  
RT "Molecular analyses of glucosyltransferase genes among strains of  
RT streptococcus mutans.";  
RL FEMS MICROBIOL. LETT. 161:331-336(1998).  
DR EMBL: D88658; D1027050; -.  
KW TRANSFERASE.  
SQ SEQUENCE 1455 AA; 163046 MW; 7659CFBC CRC32;

Query Match 91.9%; Score 136; DB 2; Length 1455;  
Best Local Similarity 95.2%; Pred. No. 7.09e-16;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

DB 507 ANDHLSILEAMSDNDTPYLHD 527  
|||||  
QY 1 ANDHLSILEAMSDNDTPYLHD 21

RESULT 10  
ID 069382 PRELIMINARY; PRT; 1455 AA.  
AC 069382;  
DT 01-AUG-1998 (TREMBLREL. 07, CREATED)  
DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)  
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
DE GLUCOSYLTRANSFERASE-SI.  
GN GTFB.  
OS STREPTOCOCCUS MUTANS.  
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; STREPTOCOCCACEAE;  
OC STREPTOCOCCUS.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN-MT8148;  
RX MEDLINE: 98231643.  
RA FUJIWARA T., TERAO Y., HOSHINO T., KAWABATA S., OOSHIMA T., SOBUE S.,  
RA KIMURA S., HAMADA S.;  
RT "Molecular analyses of glucosyltransferase genes among strains of  
RT streptococcus mutans.";  
RL FEMS MICROBIOL. LETT. 161:331-336(1998).  
DR EMBL: D88652; D1027042; -.  
KW TRANSFERASE.  
SQ SEQUENCE 1455 AA; 162970 MW; 40B022BD CRC32;

Query Match 91.9%; Score 136; DB 2; Length 1455;  
Best Local Similarity 95.2%; Pred. No. 7.09e-16;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

DB 507 ANDHLSILEAMSDNDTPYLHD 527  
|||||  
QY 1 ANDHLSILEAMSDNDTPYLHD 21

RESULT 11  
ID 055263 PRELIMINARY; PRT; 1590 AA.  
AC 055263;  
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)  
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)

DE GLUCOSYLTRANSFERASE GTF-I.  
 OS STREPTOCOCCUS SOBRINUS.  
 OC BACTERIA: FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; STREPTOCOCCACEAE;  
 STREPTOCOCCUS.  
 RN [1]  
 RN SEQUENCE FROM N.A.  
 RC STRAIN-ATCC 33478;  
 RA SATO S.;  
 RL ANN. KAGOSHIMA UNIV. DENTAL SCHOOL 16:23-29(1996).  
 DR EMBL: D63570; D1010438;  
 DR PFAM: PF00128; alpha-amylase; 1.  
 KM TRANSFERASE.  
 SQ SEQUENCE 1590 AA; 176057 MW; 4718656A CRC32;

Query Match 88.5%; Score 131; DB 2; Length 1590;  
 Best Local Similarity 85.7%; Pred. No. 1,11e-14;  
 Matches 18; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

477 ANNHVSIVEAWSNDTPYLHD 497  
 1 ANDHLSILEAWSNDTPYLHD 21

RESULT 12  
 ID 059983; PRELIMINARY; PRT; 1590 AA.  
 AC 059983;  
 DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
 DT 01-JAN-1999 (TREMBLREL. 09, LAST ANNOTATION UPDATE)  
 DE GLUCOSYLTRANSFERASE-1 PRECURSOR (EC 2.4.1.5) (DETRANSGRASE) (SUCROSE  
 6-GLUCOSYLTRANSFERASE).  
 GN GTF.  
 OS STREPTOCOCCUS SOBRINUS.  
 OC BACTERIA: FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; STREPTOCOCCACEAE;  
 STREPTOCOCCUS.  
 RN [1]  
 RN SEQUENCE FROM N.A.  
 RC STRAIN-OMZ176;  
 RX MEDLINE: 94146405.  
 RA SATO S., INOUE M., HANDA N., AIZAWA Y., ISOBE Y., KATAYAMA T.;  
 RT "DNA sequence of the glucosyltransferase gene of serotype d  
 Streptococcus sobrinus.";  
 RT DNA SEQ. 4:19-27(1993).  
 CC -1- CATALYTIC ACTIVITY: SUCROSE + (1,6-ALPHA-D-GLUCOSYL)(N) -  
 D-FRUCTOSE + (1,6-ALPHA-D-GLUCOSYL)(N+1).  
 CC EMBL: D13858; D1003482;  
 DR PFAM: PF00128; alpha-amylase; 1.  
 H SIGNAL: TRANSFERASE; GLUCOSYLTRANSFERASE.  
 FT SIGNAL 1  
 FT CHAIN 39 1590 GLUCOSYLTRANSFERASE-1.  
 SQ SEQUENCE 1590 AA; 175955 MW; F65A57D7 CRC32;

Query Match 88.5%; Score 131; DB 2; Length 1590;  
 Best Local Similarity 85.7%; Pred. No. 1,11e-14;  
 Matches 18; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 477 ANNHVSIVEAWSNDTPYLHD 497  
 1 ANDHLSILEAWSNDTPYLHD 21

QY 1 ANDHLSILEAWSNDTPYLHD 21

RESULT 13  
 ID 069398; PRELIMINARY; PRT; 1462 AA.  
 AC 069398;  
 DT 01-AUG-1998 (TREMBLREL. 07, CREATED)  
 DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
 DE GLUCOSYLTRANSFERASE-S.  
 GN GTFD.  
 OS STREPTOCOCCUS MUTANS.  
 OC BACTERIA: FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; STREPTOCOCCACEAE;  
 STREPTOCOCCUS.  
 RN [1]

RP SEQUENCE FROM N.A.  
 RC STRAIN-MT4467;  
 RX MEDLINE: 98231643.  
 RA FUJIWARA T., TERAO Y., HOSHINO T., KAWABATA S., OOSHIMA T., SOBE S.,  
 RA KIMURA S., HAMADA S.;  
 RT "Molecular analyses of glucosyltransferase genes among strains of  
 Streptococcus mutans.";  
 RT FEMS MICROBIOL. LETT. 161:331-336(1998).  
 DR EMBL: D89979; D1027061;  
 KM TRANSFERASE.  
 SQ SEQUENCE 1462 AA; 163267 MW; E9BD264F CRC32;

Query Match 68.2%; Score 101; DB 2; Length 1462;  
 Best Local Similarity 77.8%; Pred. No. 8.76e-08;  
 Matches 14; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Db 495 AINHLSILEAWSNDTPY 512  
 1 ANDHLSILEAWSNDTPY 18

QY 1 ANDHLSILEAWSNDTPY 18

RESULT 14  
 ID 069386; PRELIMINARY; PRT; 1462 AA.  
 AC 069386;  
 DT 01-AUG-1998 (TREMBLREL. 07, CREATED)  
 DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
 DE GLUCOSYLTRANSFERASE-S.  
 GN GTFD.  
 OS STREPTOCOCCUS MUTANS.  
 OC BACTERIA: FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; STREPTOCOCCACEAE;  
 STREPTOCOCCUS.  
 RN [1]  
 RN SEQUENCE FROM N.A.  
 RC STRAIN-MT4245;  
 RX MEDLINE: 98231643.  
 RA FUJIWARA T., TERAO Y., HOSHINO T., KAWABATA S., OOSHIMA T., SOBE S.,  
 RA KIMURA S., HAMADA S.;  
 RT "Molecular analyses of glucosyltransferase genes among strains of  
 Streptococcus mutans.";  
 RT FEMS MICROBIOL. LETT. 161:331-336(1998).  
 DR EMBL: D88656; D1027047;  
 KM TRANSFERASE.  
 SQ SEQUENCE 1462 AA; 163464 MW; F37E5F30 CRC32;

Query Match 68.2%; Score 101; DB 2; Length 1462;  
 Best Local Similarity 77.8%; Pred. No. 8.76e-08;  
 Matches 14; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Db 495 AINHLSILEAWSNDTPY 512  
 1 ANDHLSILEAWSNDTPY 18

QY 1 ANDHLSILEAWSNDTPY 18

RESULT 15  
 ID 069389; PRELIMINARY; PRT; 1462 AA.  
 AC 069389;  
 DT 01-AUG-1998 (TREMBLREL. 07, CREATED)  
 DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
 DE GLUCOSYLTRANSFERASE-S.  
 GN GTFD.  
 OS STREPTOCOCCUS MUTANS.  
 OC BACTERIA: FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; STREPTOCOCCACEAE;  
 STREPTOCOCCUS.  
 RN [1]  
 RN SEQUENCE FROM N.A.  
 RC STRAIN-MT4239;  
 RX MEDLINE: 98231643.  
 RA FUJIWARA T., TERAO Y., HOSHINO T., KAWABATA S., OOSHIMA T., SOBE S.,  
 RA KIMURA S., HAMADA S.;  
 RT "Molecular analyses of glucosyltransferase genes among strains of  
 Streptococcus mutans.";



RL FEMS MICROBIOL. LETT. 161:331-336(1998).  
DR EMBL; D88659; D1027051; --;  
SQ SEQUENCE 1462 AA; 163554 MW; B057EF28 CRC32;  
Query Match 68.2%; Score 101; DB 2; Length 1462;  
Best local similarity 77.8%; Pred. No. 8,76e-08;  
Matches 14; Conservative 1; Mismatches 3; Indels 0; Gaps 0;  
Db 495 AINHLSTLEAWSNDNDPY 512  
Oy 1 ANDHLSTLEAWSNDNDPY 18  
Search completed: Tue Jan 11 15:38:39 2000  
Job time : 47 secs.

---



\*\*\*\*\*  
 W I N D O W  
 \*\*\*\*\* (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.  
 Copyright (c) 1993-1998 University of Edinburgh, U.K.  
 Distribution rights by Oxford Molecular Ltd

Msrch\_pp protein - protein database search, using Smith-Waterman algorithm  
 Run on: Tue Jan 11 15:37:26 2000; Maspar time 2.42 Seconds  
 Total output not generated. 245.606 Million cell updates/sec

Title: >US-09-290-049-1  
 Description: (1-21) from US09290049.pep  
 Perfect Score: 148  
 Sequence: 1 ANDHLSTLEAMSDNDPYLHD 21

Scoring table: PAM 150  
 Gap 15

Searched: 77977 seqs, 28268293 residues

Post-processing: Minimum Match 08  
 Listing first 45 summaries

Database: swiss-prot37  
 1:swissprot

Statistics: Mean 29.534; Variance 43.172; scale 0.684

Prod. No. is the number of results predicted by chance to have a  
 score greater than or equal to the score of the result being printed,  
 and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match Length	ID	Description	Prod. No.	
1	148	100.0	1475	1 GTFB_STRMU	GLUCOSYLTRANSFERASE-I	1.49e-18
2	136	91.9	1375	1 GTFB_STRMU	GLUCOSYLTRANSFERASE-SI	1.10e-15
3	131	88.5	1592	1 GTF2_STRDO	GLUCOSYLTRANSFERASE-I	1.66e-14
4	131	88.5	1597	1 GTF1_STRDO	GLUCOSYLTRANSFERASE-I	1.66e-14
5	102	68.9	1365	1 GTFB_STRDO	GLUCOSYLTRANSFERASE-S	9.98e-08
6	101	68.2	1430	1 GTFD_STRMU	GLUCOSYLTRANSFERASE-S	7.25e-02
7	72	48.6	330	1 G3PC_LEIME	BGLF2 PROTEIN.	7.25e-02
8	72	48.6	336	1 UL16_EBY	FATTY ACID SYNTHASE, S	1.94e+00
9	66	44.6	1885	1 PCNA_CANAL	SALICYLATE BIOSYNTHET	1.94e+00
10	64	43.2	476	1 DISA_TRIGA	PUTATIVE VENOM METALLO	1.94e+00
11	64	43.2	480	1 DISA_TRIGA	PUTATIVE VENOM METALLO	1.94e+00
12	64	43.2	638	1 KRAE_XENLA	RAV PROTO-ONCOGENE SER	1.94e+00
13	63	42.6	1131	1 PMAL_DUNBI	PLASMA MEMBRANE ATPASE	4.23e+00
14	62	41.9	323	1 KRAE_MSV36	RAV SERINE/THREONINE-P	4.23e+00
15	62	41.9	380	1 KRAE_AVIMH	MIL SERINE/THREONINE-P	4.23e+00
16	62	41.9	546	1 SYL_PYRPU	ISOLEUCYL-TRNA SYNTHET	4.23e+00
17	62	41.9	647	1 KMIU_CHICK	MIL PROTO-ONCOGENE SER	4.23e+00
18	62	41.9	648	1 KRAE_RAT	RAV PROTO-ONCOGENE SER	4.23e+00
19	62	41.9	648	1 KRAE_RAT	RAV PROTO-ONCOGENE SER	4.23e+00
20	61	41.2	181	1 SYDE_ECOLI	SYD PROTEIN.	6.21e+00
21	61	41.2	202	1 HR2A_TRIPL	HEMORRHAGIC METALLOPRO	6.21e+00
22	61	41.2	330	1 G3PC_TRYB	GLYCERALDEHYDE 3-PHOSP	6.21e+00
23	61	41.2	330	1 G3PI_ECOLI	GLYCERALDEHYDE 3-PHOSP	6.21e+00

24	61	41.2	481	1 PHR_HALHA	DEOXYRIBODIPYRIMIDINE	6.21e+00
25	60	40.5	116	1 Y960_HAEIN	HYPOCHROME B6-F COMPLE	9.07e+00
26	60	40.5	180	1 UCRI_CHILT	CYTOCHROME B6-F COMPLE	9.07e+00
27	60	40.5	210	1 MAT2_YEAS	MATING-TYPE PROTEIN AL	9.07e+00
28	60	40.5	221	1 YMF2_ECOLI	HYPOHETICAL 25.1 KD P	9.07e+00
29	60	40.5	337	1 G3P_MYCE	GLYCERALDEHYDE 3-PHOSP	9.07e+00
30	60	40.5	342	1 HXCA_MOUSE	HOMEOBOX PROTEIN HOX-C	9.07e+00
31	60	40.5	471	1 YON2_YEAS	HYPOTHETICAL 51.9 KD P	9.07e+00
32	60	40.5	473	1 SYTM_SCHPO	PROBABLE THREONYL-TRNA	9.07e+00
33	60	40.5	550	1 CHIT_NPOVP	PROBABLE ENDOCHITINASE	9.07e+00
34	60	40.5	801	1 VP34_SCHPO	PHOSPHATIDYLINOSITOL 3	9.07e+00
35	60	40.5	1177	1 JAK_DROME	THROSTINE-PROTEIN KINAS	9.07e+00
36	59	39.9	158	1 GP52_BPSPI	PUTATIVE GENE PROTEIN	1.32e+01
37	59	39.9	193	1 GVE2_HALSA	GYPE PROTEIN, CHROMOSO	1.32e+01
38	59	39.9	272	1 YECE_ECOLI	HYPOTHETICAL 31.5 KD P	1.32e+01
39	59	39.9	427	1 IRE3_HUMAN	INTERFERON REGULATORY	1.32e+01
40	59	39.9	576	1 PMPI_YEAS	PERIODIC TRYPTOPHAN PR	1.32e+01
41	59	39.9	968	1 CC28_SCHPO	PUTATIVE ATP-DEPENDENT	1.32e+01
42	58	39.2	320	1 G3P_BOVIN	GLYCERALDEHYDE 3-PHOSP	1.91e+01
43	58	39.2	332	1 G3P_BUCAP	GLYCERALDEHYDE 3-PHOSP	1.91e+01
44	58	39.2	332	1 G3P_PIG	GLYCERALDEHYDE 3-PHOSP	1.91e+01
45	58	39.2	3770	1 ACVS_EMENT	DELTA-(L-ALPHA-AMINOAD	1.91e+01

## ALIGNMENTS

RESULT 1	ID	STANDARD	PRT: 1475 AA.
AC	01-NOV-1988 (REL. 09, CREATED)		
DT	01-NOV-1988 (REL. 09, LAST SEQUENCE UPDATE)		
DT	15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)		
DE	GLUCOSYLTRANSFERASE-I PRECURSOR (EC 2.4.1.5) (GTF-I) (DEXTRANSUCRASE)		
DE	(SDROSE 6-GLUCOSYLTRANSFERASE).		
GN	GTFB.		
OS	STREPTOCOCCUS MUTANS.		
OC	BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; STREPTOCOCCACEAE;		
OC	STREPTOCOCCUS.		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-GS-5;		
RX	MEDLINE: 87308013.		
RA	SHINOZA T., UEDA S., KURAMITSU H.K.;		
RT	"Sequence analysis of the gltf gene from Streptococcus mutans.";		
RL	J. BACTERIOL. 169:4263-4270(1987).		
CC	- FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS, THAT ARE THOUGHT		
CC	TO PLAY A KEY ROLE IN THE DEVELOPMENT OF THE DENTAL PLAQUE BECAUSE		
CC	OF THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDIANE THE		
CC	AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS.		
CC	- CATALYTIC ACTIVITY: SUCROSE + (1,6-ALPHA-D-GLUCOSYL) (N+1).		
CC	- D-FRUCTOSE + (1,6-ALPHA-D-GLUCOSYL) (N+1).		
CC	- SUBCELLULAR LOCATION: SECRETED.		
CC	- DISEASE: DENTAL CARIES.		
CC	- GTF-I SYNTHESIZES WATER-INSOLUBLE GLUCANS (ALPHA 1,3-LINKED		
CC	GLUCOSE AND SOME 1,6 LINKAGES), GTF-S SYNTHESIZES WATER-SOLUBLE		
CC	GLUCANS (ALPHA 1,6-GLUCOSE). GTF-SI SYNTHESIZES BOTH FORMS OF		
CC	GLUCANS.		
CC	- SIMILARITY: TO OTHER GLUCOSYLTRANSFERASES AND SOME TO A GLUCAN-		
CC	BINDING PROTEIN FROM S. MUTANS.		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (see <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>		
CC	or send an email to <a href="mailto:license@sib-sib.ch">license@sib-sib.ch</a> ).		
CC	-----		
DR	EMBL: M17361; G153640; -		
DR	PIR: B33135; B33135.		
DR	PFAM: PF00128; alpha-amylase; 1.		
KW	TRANSFERASE; GLUCOSYLTRANSFERASE; SIGNAL; REPEAT; DENTAL CARIES.		
FT	SIGNAL 1 34		POTENTIAL.

```

FT CHAIN 35 1475 GLUCOSYLTRANSFERASE-1.
FT DOMAIN 35 1050 CATALYTIC (APPROXIMATE).
FT TRANSFERASE 1475 1475 GLUCAN-BINDING (APPROXIMATE).
FT REPEAT 1096 1129 A REPEAT.
FT DOMAIN 1160 1469 5 X TANDEM REPEATS.
FT REPEAT 1160 1209 1.
FT REPEAT 1224 1274 2.
FT REPEAT 1289 1339 3.
FT REPEAT 1354 1404 4.
FT REPEAT 1419 1469 5.
FT REPEAT 1475 165812 MM; 4542C1D6 CRC32;
SQ SEQUENCE 1475 AA; 165812 MM; 4542C1D6 CRC32;

Query Match 100.0%; Score 148; DB 1; Length 1475;
Best Local Similarity 100.0%; Pred. No. 1,49e-18;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 481 ANDHLSILEMSNDPTPYLHD 501
1 ANDHLSILEMSNDPTPYLHD 21

RESULT 2
ID GTF2_STRMU STANDARD; PRT; 1375 AA.
AC P13470; P05427;
DT 01-NOV-1988 (REL. 09, CREATED)
DT 01-JAN-1990 (REL. 13, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE GLUCOSYLTRANSFERASE-SI PRECURSOR (EC 2.4.1.5) (GTF-SI)
DE (DEXTRANSUCRASE) (SUCROSE 6-GLUCOSYLTRANSFERASE).
GN GTF.
OS STREPTOCOCCUS MUTANS.
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; STREPTOCOCCACEAE;
OC STREPTOCOCCUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-GS-5;
RX MEDLINE; 89137980.
RA UEDA S., SHIROZA T., KURAMITSU H.K.;
RT "Sequence analysis of the gtfC gene from Streptococcus mutans GS-5.";
RL GENE 69:101-109(1988).
RN [2]
RP SEQUENCE OF 1-349 FROM N.A.
RC STRAIN-GS-5;
RX MEDLINE; 87308013.
RA SHIROZA T., UEDA S., KURAMITSU H.K.;
RT "Sequence analysis of the gtfB gene from Streptococcus mutans.";
RL J. BACTERIO. 169:4263-4270(1987).
CC -1- FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS, THAT ARE THOUGHT
TO PLAY A KEY ROLE IN THE DEVELOPMENT OF THE DENTAL PLAQUE BECAUSE
OF THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDIATE THE
AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS.
CC -1- CATALYTIC ACTIVITY: SUCROSE + (1,6-ALPHA-D-GLUCOSYL) (N) -
D-FRUCTOSE + (1,6-ALPHA-D-GLUCOSYL) (N+1).
CC -1- SUBCELLULAR LOCATION: SECRETED.
CC -1- DISEASE: DENTAL CARIES.
CC -1- GTF-1 SYNTHESIZES WATER-INSOLUBLE GLUCANS (ALPHA 1,3-LINKED
GLUCOSE AND SOME 1,6 LINKAGES), GTF-S SYNTHESIZES WATER-SOLUBLE
GLUCANS (ALPHA 1,6-GLUCOSE). GTF-SI SYNTHESIZES BOTH FORMS OF
GLUCANS.
CC -1- SIMILARITY: TO OTHER GLUCOSYLTRANSFERASES AND SOME TO A GLUCAN-
BINDING PROTEIN FROM S. MUTANS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M22054; G153643; -
CC DR EMBL: M17361; G153641; -
CC PIR: J70345; J70345.

```

```

DR PIR: C33135; C33135.
DR PFAM: PF00128; alpha-amylase; 1.
KW TRANSFERASE; GLYCOSYLTRANSFERASE; SIGNAL; REPEAT; DENTAL CARIES.
FT SIGNAL 1 34
FT CHAIN 35 1375 GLUCOSYLTRANSFERASE-SI.
FT DOMAIN 35 1050 CATALYTIC (APPROXIMATE).
FT DOMAIN 1126 1375 GLUCAN-BINDING (APPROXIMATE).
FT REPEAT 1126 1159 2,4 A, 1 C AND 1 AC REPEATS.
FT REPEAT 1169 1200 A REPEAT.
FT REPEAT 1227 1238 C REPEAT.
FT REPEAT 1253 1303 AC REPEAT.
FT REPEAT 1318 1330 A REPEAT (INCOMPLETE).
FT REPEAT 1375 153022 MM; 3EA3727E CRC32;
SQ SEQUENCE 1375 AA; 153022 MM; 3EA3727E CRC32;

Query Match 91.9%; Score 136; DB 1; Length 1375;
Best Local Similarity 95.2%; Pred. No. 1,10e-15;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 507 ANDHLSILEMSNDPTPYLHD 527
1 ANDHLSILEMSNDPTPYLHD 21

OY 1 ANDHLSILEMSNDPTPYLHD 21

RESULT 3
ID GTF2_STRDO STANDARD; PRT; 1592 AA.
AC P27470;
DT 01-AUG-1992 (REL. 23, CREATED)
DT 01-AUG-1992 (REL. 23, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE GLUCOSYLTRANSFERASE-1 PRECURSOR (EC 2.4.1.5) (GTF-1) (DEXTRANSUCRASE)
DE (SUCROSE 6-GLUCOSYLTRANSFERASE).
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; STREPTOCOCCACEAE;
OC STREPTOCOCCUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-6715;
RX MEDLINE; 91123227.
RA ABO H., MATSUMURA T., KODAMA T., OHTA H., FUKUI K., KATO K.,
RA KAGAWA H.;
RT "Peptide sequences for sucrose splitting and glucan binding within
RT Streptococcus sobrinus glucosyltransferase (water-insoluble glucan
RT synthetase).";
RL J. BACTERIO. 173:989-996(1991).
CC -1- FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS, THAT ARE THOUGHT
TO PLAY A KEY ROLE IN THE DEVELOPMENT OF THE DENTAL PLAQUE BECAUSE
OF THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDIATE THE
AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS.
CC -1- CATALYTIC ACTIVITY: SUCROSE + (1,6-ALPHA-D-GLUCOSYL) (N) -
D-FRUCTOSE + (1,6-ALPHA-D-GLUCOSYL) (N+1).
CC -1- SUBCELLULAR LOCATION: SECRETED.
CC -1- DISEASE: DENTAL CARIES.
CC -1- GTF-1 SYNTHESIZES WATER-INSOLUBLE GLUCANS (ALPHA 1,3-LINKED
GLUCOSE AND SOME 1,6 LINKAGES), GTF-S SYNTHESIZES WATER-SOLUBLE
GLUCANS (ALPHA 1,6-GLUCOSE). GTF-SI SYNTHESIZES BOTH FORMS OF
GLUCANS.
CC -1- SIMILARITY: TO OTHER GLUCOSYLTRANSFERASES AND SOME TO A GLUCAN-
BINDING PROTEIN FROM S. MUTANS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: D90213; G217033; -
CC DR EMBL: A38175; A38175.
CC DR PFAM: PF00128; alpha-amylase; 1.
CC DR HSP: P00695; ZHEB.
CC TRANSFERASE; GLYCOSYLTRANSFERASE; SIGNAL; REPEAT; DENTAL CARIES.
KW

```

```

FT SIGNAL 1 38 POTENTIAL.
FT CHAIN 39 1592 GLUCOSYLTRANSFERASE-I.
FT DOMAIN 39 1044 CATALYTIC (APPROXIMATE).
FT DOMAIN 1093 1592 GLUCAN-BINDING (APPROXIMATE).
FT DOMAIN 1093 1592 6.5 X TANDEM REPEATS.
FT REPEAT 1093 1142 1.
FT REPEAT 1158 1207 2.
FT REPEAT 1222 1272 3.
FT REPEAT 1287 1337 4.
FT REPEAT 1402 1451 5.
FT REPEAT 1514 1563 6.
FT REPEAT 1577 1592 7 (INCOMPLETE).
SQ SEQUENCE 1592 AA; 176167 MW; 273A23FA CRC32;

Query Match
Best Local Similarity 85.7%; Score 131; DB 1; Length 1592;
Matches 18; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 477 ANNHVSIIVASDNDPTPYLHD 497
1 ANDHLSILEAWSNDPTPYLHD 21

RESULT 4
ID GTF1_STRDO STANDARD; PRT; 1597 AA.
AC P11001;
DT 01-JUL-1989 (REL. 11, CREATED)
DT 01-JUL-1989 (REL. 11, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE GLUCOSYLTRANSFERASE-1 PRECURSOR (EC 2.4.1.5) (GTF-1) (DEXTRANSUCRASE)
DE (SUCROSE 6-GLUCOSYLTRANSFERASE).
GN GTF1.
OS STREPTOCOCCUS DOWNEI (STREPTOCOCCUS SOBRINUS).
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; STREPTOCOCCACEAE;
OC STREPTOCOCCUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 87308014.
RA FERRETTI J.J., GILPIN M.L., RUSSELL R.R.B.;
RT "Nucleotide sequence of a glucosyltransferase gene from Streptococcus
sobiinus Me28."
RL J. BACTERIOL. 169:4271-4278(1987).
CC -1- FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS, THAT ARE THOUGHT
CC TO PLAY A KEY ROLE IN THE DEVELOPMENT OF THE DENTAL PLAQUE BECAUSE
CC OF THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDIATE THE
CC AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS.
CC -1- CATALYTIC ACTIVITY: SUCROSE + (1,6-ALPHA-D-GLUCOSYL)(N) -
CC -1- D-FRUCTOSE + (1,6-ALPHA-D-GLUCOSYL)(N+1).
CC -1- SUBCELLULAR LOCATION: SECRETED.
CC -1- DISEASE: DENTAL CARIES.
CC -1- GTF-1 SYNTHESIZES WATER-INSOLUBLE GLUCANS (ALPHA 1,3-LINKED
CC GLUCANS AND SOME 1,6 LINKAGES), GTF-S SYNTHESIZES WATER-SOLUBLE
CC GLUCANS (ALPHA 1,6-GLUCOSE). GTF-SI SYNTHESIZES BOTH FORMS OF
CC GLUCANS.
CC -1- SIMILARITY: TO OTHER GLUCOSYLTRANSFERASES AND SOME TO A GLUCAN-
CC BINDING PROTEIN FROM S. MUTANS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M17391; G153647; -
CC PFAM; PF00128; alpha-amylase; 1.
CC TRANSFERASE; GLUCOSYLTRANSFERASE; SIGNAL; REPEAT; DENTAL CARIES.
CC SIGNAL 1 38
CC CHAIN 39 1597 GLUCOSYLTRANSFERASE-1.
CC DOMAIN 39 1050 CATALYTIC (APPROXIMATE).
CC DOMAIN 1099 1597 GLUCAN-BINDING (APPROXIMATE).

```

```

FT DOMAIN 1099 1597 1.25 A, 2 B, AND 5 AC REPEATS.
FT REPEAT 1099 1132 A REPEAT.
FT REPEAT 1163 1213 AC REPEAT.
FT REPEAT 1227 1277 AC REPEAT.
FT REPEAT 1292 1342 AC REPEAT.
FT REPEAT 1352 1399 B REPEAT.
FT REPEAT 1406 1455 AC REPEAT.
FT REPEAT 1465 1512 B REPEAT.
FT REPEAT 1519 1568 AC REPEAT.
FT REPEAT 1582 1597 A REPEAT (INCOMPLETE).
SQ SEQUENCE 1597 AA; 177080 MW; 995233CA CRC32;

Query Match
Best Local Similarity 85.7%; Score 131; DB 1; Length 1597;
Matches 18; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 483 ANNHVSIIVASDNDPTPYLHD 503
1 ANDHLSILEAWSNDPTPYLHD 21

RESULT 5
ID GTF5_STRDO STANDARD; PRT; 1365 AA.
AC P29336;
DT 01-DEC-1992 (REL. 24, CREATED)
DT 01-DEC-1992 (REL. 24, LAST SEQUENCE UPDATE)
DT 01-APR-1993 (REL. 25, LAST ANNOTATION UPDATE)
DE GLUCOSYLTRANSFERASE-5 PRECURSOR (EC 2.4.1.5) (GTF-S) (DEXTRANSUCRASE)
DE (SUCROSE 6-GLUCOSYLTRANSFERASE).
GN GTF5.
OS STREPTOCOCCUS DOWNEI (STREPTOCOCCUS SOBRINUS).
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; STREPTOCOCCACEAE;
OC STREPTOCOCCUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 90316665.
RA GILMORE K.S., RUSSELL R.R., FERRETTI J.J.;
RT "Analysis of the Streptococcus downei gtf5 gene, which specifies a
glucosyltransferase that synthesizes soluble glucans."
RL INFECT. IMMUN. 58:2452-2458(1990).
CC -1- FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS, THAT ARE THOUGHT
CC TO PLAY A KEY ROLE IN THE DEVELOPMENT OF THE DENTAL PLAQUE BECAUSE
CC OF THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDIATE THE
CC AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS.
CC -1- CATALYTIC ACTIVITY: SUCROSE + (1,6-ALPHA-D-GLUCOSYL)(N) -
CC -1- D-FRUCTOSE + (1,6-ALPHA-D-GLUCOSYL)(N+1).
CC -1- ENZYME REGULATION: GLUCAN SYNTHESIS BY GTF-S IS INDEPENDENT OF
CC PRIMER GLUCAN UNLIKE GTF-1.
CC -1- DISEASE: DENTAL CARIES.
CC -1- GTF-S SYNTHESIZES WATER-SOLUBLE GLUCANS (ALPHA 1,6-GLUCOSE).
CC -1- SIMILARITY: TO OTHER GLUCOSYLTRANSFERASES AND SOME TO A GLUCAN-
CC BINDING PROTEIN FROM S. MUTANS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M30943; G153653; -
CC PIR; A41483; A41483.
CC PFAM; PF00128; alpha-amylase; 1.
CC TRANSFERASE; GLUCOSYLTRANSFERASE; SIGNAL; REPEAT; DENTAL CARIES.
CC SIGNAL 1 36
CC CHAIN 37 1365
CC DOMAIN 37 1050 CATALYTIC (APPROXIMATE).
CC DOMAIN 1083 1365 GLUCAN-BINDING (APPROXIMATE).
CC DOMAIN 1083 1365 4.5 X TANDEM REPEATS.
CC REPEAT 1083 1131 1.
CC REPEAT 1150 1199 2.

```

FT REPEAT 1225 1274 3.  
FT REPEAT 1289 1339 4.  
FT REPEAT 1353 1365 5 (INCOMPLETE).  
SQ SEQUENCE 1365 AA; 151590 MW; 156F8B31 CRC32;

Query Match 68.9%; Score 102; DB 1; Length 1365;  
Best Local Similarity 71.4%; Pred. No. 6,05e-08;  
Matches 15; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

DB 467 AIDHLSILEAMSGNDNDYK 487  
1 AIDHLSILEAMSGNDNDYK 21

RESULT 6  
ID GTFD\_STRMU STANDARD; PRT; 1430 AA.  
AC P49331;  
DT 01-FEB-1996 (REL. 33, CREATED)  
01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)  
01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
DE GLUCOSYLTRANSFERASE-S PRECURSOR (EC 2.4.1.5) (GTF-S) (DEXTRANSUCRASE)  
DE (SUCROSE 6-GLUCOSYLTRANSFERASE).  
GN GTFD.  
OS STREPTOCOCCUS MUTANS.  
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; STREPTOCOCCACEAE;  
OC STREPTOCOCCUS.  
RN (1)  
RP SEQUENCE FROM N.A.  
RC STRAIN-GS-5;  
RX MEDLINE: 91100958.  
RA HONDA O., KATO C., KURAMITSU H.K.;  
RT "Nucleotide sequence of the Streptococcus mutans gtfD gene encoding the glucosyltransferase-S enzyme."  
RL J. GEN. MICROBIOL. 136:2099-2105(1990).  
CC -1- FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS, THAT ARE THOUGHT TO PLAY A KEY ROLE IN THE DEVELOPMENT OF THE DENTAL PLAQUE BECAUSE OF THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDIANE THE AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS.  
CC -1- CATALYTIC ACTIVITY: SUCROSE + (1,6-ALPHA-D-GLUCOSYL)(N) -  
CC -1- D-FRUCTULOSE + (1,6-ALPHA-D-GLUCOSYL)(N+1).  
CC -1- SUBCELLULAR LOCATION: SECRETED.  
CC -1- DISEASE: DENTAL CARIES.  
CC -1- GTF-1 SYNTHESIZES WATER-INSOLUBLE GLUCANS (ALPHA 1,3-LINKED GLUCOSE AND SOME 1,6 LINKAGES), GTF-S SYNTHESIZES WATER-SOLUBLE GLUCANS (ALPHA 1,6-GLUCOSE). GTF-SI SYNTHESIZES BOTH FORMS OF GLUCANS.  
CC -1- SIMILARITY: TO OTHER GLUCOSYLTRANSFERASES AND SOME TO A GLUCAN-BINDING PROTEIN FROM S. MUTANS.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC EMBL: M29296; G153645; -  
DR PFAM: PF00128; alpha-amylase; 1.  
KM TRANSFERASE; GLUCOSYLTRANSFERASE; SIGNAL; REPEAT; DENTAL CARIES.  
FT SIGNAL 1 2 POTENTIAL.  
FT CHAIN 1 1430  
FT DOMAIN 1232 1423 3 X 63 AA APPROXIMATE TANDEM REPEATS.  
FT REPEAT 1232 1295 1.  
FT REPEAT 1296 1359 2.  
FT REPEAT 1360 1423 3.  
SQ SEQUENCE 1430 AA; 159765 MW; D3DE3681 CRC32;

Query Match 68.2%; Score 101; DB 1; Length 1430;  
Best Local Similarity 77.8%; Pred. No. 9,98e-08;  
Matches 14; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

DB 495 AINHLSILEAMSGNDNDY 512

QY 1 AIDHLSILEAMSGNDNDY 18  
1 AIDHLSILEAMSGNDNDY 18

RESULT 7  
ID G3PC\_LETME STANDARD; PRT; 330 AA.  
AC 001558;  
DT 01-APR-1993 (REL. 25, CREATED)  
01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)  
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)  
DE GLYCERALDEHYDE 3-PHOSPHATE DEHYDROGENASE, CYTOSOLIC (EC 1.2.1.12) (GAPDH).  
GN GAPDH.  
OS LEISHMANIA MEXICANA.  
OC EUKARYOTA; EUKLENOZOA; KINETOPLASTIDA; TRYPAANOMATIDAE; LEISHMANIA.  
RN (1)  
RP SEQUENCE FROM N.A.  
RC STRAIN-SSP. MEXICANA;  
RX MEDLINE: 93063042.  
RA HANNAERT V., BLAAUW M., KOHL L., ALLERT S., OPPERDOES F.R.,  
RA MICHELIS P.A.M.;  
RT "Molecular analysis of the cytosolic and glycosomal glyceraldehyde-3-phosphate dehydrogenase in Leishmania mexicana.";  
RL MOL. BIOCHEM. PARASITOL. 55:115-126(1992).  
CC -1- CATALYTIC ACTIVITY: D-GLYCERALDEHYDE 3-PHOSPHATE + ORTHOPHOSPHATE + NAD(+) = 1,3-DIPHOSPHATEGLYCERATE + NADH.  
CC -1- PATHWAY: FIRST STEP IN THE SECOND PHASE OF GLYCOLYSIS.  
CC -1- SUBUNIT: HOMOTETRAMER.  
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC EMBL: X65220; G9553; -  
DR PIR: S25142; S25142.  
DR PIR: B48445; B48445.  
DR PROSITE: PS00071; GAPDH; 1.  
DR PFAM: PF00044; gpdh; 1.  
DR HSSP: P06977; IGAD.  
KM GLYCOLYSIS; OXIDOREDUCTASE; NAD.  
FT INIT MET 0 0  
FT BINDING 148 148  
FT ACT\_SITE 175 175  
SQ SEQUENCE 330 AA; 35511 MW; 80B6A7E9 CRC32;

Query Match 48.6%; Score 72; DB 1; Length 330;  
Best Local Similarity 52.6%; Pred. No. 7,25e-02;  
Matches 10; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

DB 300 NDHFVKLVSYNDYGTSH 318  
1 NDHFVKLVSYNDYGTSH 20

RESULT 8  
ID U1L6\_EBV STANDARD; PRT; 336 AA.  
AC P03221;  
DT 21-JUL-1986 (REL. 01, CREATED)  
21-JUL-1986 (REL. 01, LAST SEQUENCE UPDATE)  
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)  
DE BGLF2 PROTEIN.  
GN BGLF2.  
OS EPSTEIN-BARR VIRUS (STRAIN B95-8) (HUMAN HERPESVIRUS 4).  
OS VIRUSES; DSDNA VIRUSES, NO RNA STAGE; HERPESVIRIDAE;  
OC GAMMAHERPESVIRINAE; LYMPHOCRYPTOVIRUS.  
RN (1)  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 84270667.

```

CC      TWO MULTIFUNCTIONAL SUBUNITS (ALPHA & BETA).
CC      -1- SIMILARITY: TO THE FATTY ACID SYNTHETASE, SUBUNIT ALPHA FROM
CC      OTHER FUNGI.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
CC      EMBL; L29063; G456443; -
CC      DR PROSITE; PS00012; PHOSPHOPANTETHEINE; 1.
CC      DR PROSITE; PS00605; B-KETOACYL SYNTHASE; 1.
CC      KM FATTY ACID BIOSYNTHESIS; MULTIFUNCTIONAL ENZYME; OXIDOREDUCTASE;
CC      TRANSFERASE; NADP; PHOSPHOPANTETHEINE.
CC      FT DOMAIN 1 ? ? ACYL CARRIER.
CC      FT DOMAIN ? ? BETA-KETOACYL REDUCTASE.
CC      FT BINDING ? 1885 BETA-KETOACYL SYNTHASE.
CC      FT ACT_SITE 1304 1304 PHOSPHOPANTETHEINE (BY SIMILARITY).
CC      SQ SEQUENCE 1885 AA; 207588 MW; A19558C4 CRC32;
CC      -----
CC      Query Match 44.6%; Score 66; DB 1; Length 1885;
CC      Best Local Similarity 25.0%; Pred. No. 8,72e-01;
CC      Matches 5; Conservative 8; Mismatches 7; Indels 0; Gaps 0;
CC      -----
CC      Db 1452 NRRLEQIKSWETELSYLQE 1471
CC      Oy 2 NDHLSTLEWSDNDPTPLMD 21
CC      1 : 1 : : : : : : :
CC      ID PCHA_PSEAE STANDARD; PRT; 476 AA.
CC      AC Q51508;
CC      DT 01-NOV-1997 (REL. 35, CREATED)
CC      DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
CC      DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
CC      DE SALICYLATE BIOSYNTHESIS ISOCHORISMATE SYNTHASE (EC 5.4.99.6).
CC      GN PCHA.
CC      OS PSEUDOMONAS AERUGINOSA.
CC      CC BACTERIA: PROTEOBACTERIA; GAMMA SUBDIVISION; PSEUDOMONAS GROUP;
CC      CC PSEUDOMONAS.
CC      RN [1]
CC      RP SEQUENCE FROM N.A.
CC      RC STRAIN-ATCC 15692 / PAOI;
CC      RX MEDLINE; 96086939.
CC      RA SERINO L., REIMANN C., BAUR H., BEYELER M., VISCA P., HAAS D.;
CC      RT "Structural genes for salicylate biosynthesis from chorismate in
CC      Pseudomonas aeruginosa."
CC      RL MOL. GEN. GENET. 249:217-228(1995).
CC      CC -1- FUNCTION: INVOLVED IN THE CONVERSION OF CHORISMATE TO SALICYLATE
CC      (PROBABLY).
CC      CC -1- CATALYTIC ACTIVITY: CHORISMATE - ISOCHORISMATE.
CC      CC -1- PATHWAY: SALICYLATE BIOSYNTHESIS.
CC      CC -1- SIMILARITY: STRONG, TO OTHER ISOCHORISMATE SYNTHASES; WEAK, TO
CC      TYPE AND PABB.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
CC      EMBL; X82644; G929780; -
CC      DR PFAM; PF00425; chorismate_bind. 1.
CC      KM ISOMERASE.
CC      SQ SEQUENCE 476 AA; 52071 MW; DEDC2D5C CRC32;
CC      -----
CC      Query Match 43.2%; Score 64; DB 1; Length 476;

```

Best Local Similarity 47.1%; Pred. No. 1.94e+00;  
Matches 8; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

```

Db      44 DPLQVEGAWDDRQTPCL 60
      | | : | | : | | |
QY      3 DHLSTLEAWSNDNDTPYL 19

```

RESULT	11
ID	DISA_TRIGA
STANDARD;	PRT;
480 AA	

DT 01-APR-1990 (REL. 14, CREATED)  
DT 01-AUG-1992 (REL. 23, LAST SEQUENCE UPDATE)  
DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)  
DE PUTATIVE VENOM METALLOPROTEINASE PRECURSOR (EC 3.4.24.-) [CONTAINS:  
DE DISINTEGRIN TRIGRAMIN ALPHA (PLATELET AGGREGATION ACTIVATION  
DE INHIBITOR)].  
OS TRIMEESURUS GRAMINEUS (INDIAN GREEN TREE VIPER) (GREEN HAIRY SNAKE)  
OS ENKAROTIA METAZOA: CHORDATA: VERTEBRATA: LEPTOSOMATA: SOMAMATA:  
SCLEROGLOSSA: SERPENTES: COLUBROIDEA: VIPERIDAE: CROTALINAE,  
TRIMEESURUS.

```
CC -----
DR EMBL; X51530; G64408; -.
DR PIR; A29784; A29784.
DR PIR; A30065; A30065.
DR
```

Query Match	43.2%	Score 64;	DB 1;	Length 480;
Best Local Similarity	63.6%	Pred. No. 1.94e+00;		
Matches	7;	Conservative	3;	Mismatches 1;
				Indels 0;
				Gaps 0;

Db	244	LSVLEIWSKD	254
		:     :	
QY	5	LSILEAWSND	15

```

ID RESULT 12
AC AC KRAF_XENIA STANDARD; PRT; 638 AA.
AC AC P09560; 091390;
DT DT 01-MAR-1989 (REL. 10, CREATED)
DT DT 01-MAR-1989 (REL. 10, LAST SEQUENCE UPDATE)
DT DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE DE RAF_PROTO-ONCOGENE SERINE/THROMBIN-PROTEIN KINASE (BC 2.7.1.-).
GN GN C-RAF.
OS OS XENOPUS LAEVIS (AFRICAN CLAMMED FROG).
OC OC EURAROTIA, METAZOA: CHORDATA: VERTEBRATA: AMPHIBIA: BATRACHIA; ANURA
OC OC MESOBATRACHIA: PIPOIDEA, PIPIDAE; XENOPODINAE; XENOPUS.
RN RN [1]
RP RP SEQUENCE FROM N.A.
RC RC TISSUE-OOCYTE.
RX RX MEDLINE: 89057471.
RA RA LE GUELLEC R., LE GUELLEC K., PARIS J., PHILIPPE M.;
RT RT "Nucleotide sequence of Xenopus C-raf coding region.";
RL RL NUCLEIC ACIDS RES. 16:10357-10357(1988).
RN RN [2]
RP RP SEQUENCE FROM N.A.
RX RX MEDLINE: 92096753.
RA RA LE GUELLEC R., COUVRIER A., LE GUELLEC K., PARIS J., LE FUR N.,
RA RA PHILIPPE M.;
RT RT "Xenopus c-raf proto-oncogene: cloning and expression during
RT RT oogenesis and early development.";
RL RL BIOL. CELL 72:39-45(1991).
CC CC -1- FUNCTION: INVOLVED IN THE TRANSDUCTION OF MITOGENIC SIGNALS FROM
CC CC THE CELL MEMBRANE TO THE NUCLEUS, PART OF THE RAF-DEPENDENT
CC CC SIGNALING PATHWAY FROM RECEPTORS TO THE NUCLEUS.
CC CC -1- SIMILARITY: WITH THE CONSERVED CATALYTIC DOMAINS OF SER/THR-
CC CC PROTEIN KINASES, BELONGS TO THE MIL/RAF SUBFAMILY.
CC CC -1- SIMILARITY: CONTAINS A COPY OF THE ZINC-DEPENDENT PHORBOL-
CC CC ESTER AND DAG BINDING DOMAIN.

```



CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL: X12948; G65028; -  
 CC EMBL: S74063; G241260; -  
 CC PIR: S01930; TVXLRP.

DR PROSITE: PS00107; PROTEIN\_KINASE\_ATP; 1.  
 DR PROSITE: PS00108; PROTEIN\_KINASE\_ST; 1.  
 DR PROSITE: PS00479; DAG\_PE\_BINDING\_DOMAIN; 1.  
 DR PROSITE: PS50011; PROTEIN\_KINASE\_DOM; 1.  
 DR PFAM: PF00069; PKinase; 1.  
 DR PFAM: PF00130; DAG-pe-bind; 1.  
 DR HSPF: P04049; lfgi.

KW TRANSFERASE: SERINE/THREONINE-PROTEIN KINASE; PROTO-ONCOGENE; ZINC;  
 KW ATP-BINDING; PHORBOL-ESTER BINDING.  
 FT DOMAIN 138 183 PHORBOL-ESTER AND DAG BINDING.  
 FT NP\_BIND 340 600 PROTEIN KINASE.  
 FT BINDING 346 354 ATP (BY SIMILARITY).  
 FT ACT\_SITE 459 459 ATP (BY SIMILARITY).  
 FT CONFLICT 309 309 K -> R (IN REF. 2).  
 SQ SEQUENCE 638 AA; 71959 MW; A46048CC CRC32;

Query Match 43.2%; Score 64; DB 1; Length 638;  
 Best Local Similarity 30.0%; Pred. No. 1.94e+00;  
 Matches 6; Conservative 7; Mismatches 7; Indels 0; Gaps 0;

Db 404 TKDNLAVTQWCGSSLYH 423  
 1 ANDHSLILNMSDNTPLYH 20

RESULT 13  
 ID PMAL.DUNBI STANDARD; PRT; 1131 AA.  
 AC P54211;  
 DT 01-OCT-1996 (REL. 34, CREATED)  
 DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)  
 DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)  
 DE PLASMA MEMBRANE ATPASE (EC 3.6.1.35) (PROTON PUMP).  
 GN PMAL.  
 OS DUNALIELLA BIOCULATA.  
 OC EUKARYOTA; VIRIDIPHYTES; CHLOROPHYTES; VOLVOCALES;  
 OC DUNALIACEAE; DUNALIELLA.  
 [1]  
 SEQUENCE FROM N.A.  
 STRAIN-SAG 19-4;  
 MEDLINE: 95375230.  
 RA WOLF A.H., SLATMAN C.W., GRADMAN D.;  
 RT "Primary structure of the plasma membrane H(+)-ATPase from the  
 RT halotolerant alga Dunaliella bioculata.";  
 RL PLANT MOL. BIOL. 28:657-666(1995).  
 CC -1- FUNCTION: THE PLASMA MEMBRANE ATPASE OF PLANTS AND FUNGI IS A  
 CC HYDROGEN ION PUMP. THE PROTON GRADIENT IT GENERATES DRIVES THE  
 CC ACTIVE TRANSPORT OF NUTRIENTS BY H+-STIMPT. THE RESULTING  
 CC EXTERNAL ACIDIFICATION AND/OR INTERNAL ALKALIZATION MAY MEDIANTE  
 CC GROWTH RESPONSES.  
 CC -1- CATALYTIC ACTIVITY: ATP + H(2)O -> ADP + ORTHOPHOSPHATE.  
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN  
 CC -1- SIMILARITY: BELONGS TO THE CATION TRANSPORT ATPASES FAMILY  
 CC (E1-E2 ATPASES).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL: X73901; G313704; -  
 CC PROSITE: PS00154; ATPASE\_E1\_E2; 1.  
 DR PFAM: PF00122; E1-E2 ATPase; 1.  
 KW HYDROLASE; HYDROGEN ION TRANSPORT; TRANSMEMBRANE; PHOSPHORYLATION;  
 KW ATP-BINDING.  
 FT TRANSMEM 77 97 POTENTIAL.  
 FT TRANSMEM 98 118 POTENTIAL.  
 FT TRANSMEM 151 171 POTENTIAL.  
 FT TRANSMEM 231 251 POTENTIAL.  
 FT TRANSMEM 265 285 POTENTIAL.  
 FT TRANSMEM 305 325 POTENTIAL.  
 FT TRANSMEM 642 662 POTENTIAL.  
 FT TRANSMEM 689 709 POTENTIAL.  
 FT TRANSMEM 733 753 POTENTIAL.  
 FT TRANSMEM 884 904 POTENTIAL.  
 FT TRANSMEM 946 966 POTENTIAL.  
 FT MOD\_RES 357 357 PHOSPHORYLATION (PROBABLE).  
 SQ SEQUENCE 1131 AA; 123450 MW; A71E561 CRC32;

Query Match 42.6%; Score 63; DB 1; Length 1131;  
 Best Local Similarity 35.7%; Pred. No. 2.87e+00;  
 Matches 5; Conservative 7; Mismatches 2; Indels 0; Gaps 0;

Db 774 NDQSYVLQPCWDE 787  
 2 NDHSLILNMSDND 15

RESULT 14  
 ID KRAF.MSV36 STANDARD; PRT; 323 AA.  
 AC P00532;  
 DT 21-JUL-1986 (REL. 01, CREATED)  
 DT 21-JUL-1986 (REL. 01, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)  
 DE RAF SERINE/THREONINE-PROTEIN KINASE TRANSFORMING PROTEIN (EC 2.7.1.-).  
 GN V-RAF.  
 OS MURINE SARCOMA VIRUS 3611.  
 OC VIRUSES; RETROVIRUSES; RETROVIRIDAE; MAMMALIAN TYPE C RETROVIRUSES.  
 [1]  
 SEQUENCE FROM N.A.  
 MEDLINE: 84121298.  
 RA KAN N.C., FLORELLIS C.S., MARK G.E., DUESBERG P.H., PAPAS T.S.;  
 RT "A common onc gene sequence transduced by avian carcinoma virus MH2  
 RT and by murine sarcoma virus 3611.";  
 RL SCIENCE 223:813-816(1984).  
 [2]  
 SEQUENCE FROM N.A.  
 MEDLINE: 84172180.  
 RA MARK G.E., RAPP U.R.;  
 RT "Primary structure of v-raf: relatedness to the src family of  
 RT oncogenes.";  
 RL SCIENCE 224:285-289(1984).  
 CC -1- THIS PROTEIN IS SYNTHESIZED AS A GAG-RAF POLYPEPTIDE.  
 CC -1- SIMILARITY: WITH THE CONSERVED CATALYTIC DOMAINS OF SSR/THR-  
 CC PROTEIN KINASES. BELONGS TO THE MLT/RAF SUBFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL: K01691; G332207; ALT\_INIT.  
 DR PIR: A00638; TVMVP6.  
 DR PROSITE: PS00107; PROTEIN\_KINASE\_ATP; 1.  
 DR PROSITE: PS00108; PROTEIN\_KINASE\_ST; 1.  
 DR PROSITE: PS50011; PROTEIN\_KINASE\_DOM; 1.  
 DR PFAM: PF00069; PKinase; 1.  
 DR HSPF: P11362; lfgi.

KW POLYPROTEIN; TRANSFORMING PROTEIN; SERINE/THREONINE-PROTEIN KINASE;  
 KM TRANSFERASE; ONCOGENE; ATP-BINDING.  
 FT DOMAIN 24 284 PROTEIN KINASE.  
 FT NP\_BIND 30 38 ATP (BY SIMILARITY).  
 FT BINDING 50 50 ATP (BY SIMILARITY).  
 FT ACT SITE 143 143 BY SIMILARITY.  
 SQ SEQUENCE 323 AA; 36883 MW; B3FF54BE CRC32;

Query Match 41.9%; Score 62; DB 1; Length 323;  
 Best Local Similarity 30.0%; Pred. No. 4.23e+00;  
 Matches 6; Conservative 7; Mismatches 7; Indels 0; Gaps 0;

Db 88 TKDNLAIYQWCGSSLYKH 107  
 OY 1 ANDHLSLEAMSDNDPYLH 20

RESULT 15  
 KMILAVIMH STANDARD; PRT; 380 AA.

DI 21-JUL-1986 (REL. 01, CREATED)  
 DT 21-JUL-1986 (REL. 01, LAST SEQUENCE UPDATE)  
 DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)  
 DE MTL SERINE/THREONINE-PROTEIN KINASE TRANSFORMING PROTEIN  
 (EC 2.7.1.1).  
 GN V-MIL OR V-MHT.  
 OS AVIAN RETROVIRUS MH2.  
 OC VIRUSES; RETROID VIRUSES; RETROVIRIDAE; AVIAN TYPE C RETROVIRUSES.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 84191511.  
 RA SOTRAVE P., BONNER T.I., RAPP U.R., JANSEN H.W., PATSCHINSKY T.,  
 RA BISTER K.;  
 RT "Nucleotide sequence of avian retroviral oncogene v-mil: homologue of  
 RT murine retroviral oncogene v-raf.";  
 RL NATURE 309:85-88(1984).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 84121298.  
 RA KAN N.C., FLOREDELIS C.S., MARK G.E., DUESBERG P.H., PAPAS T.S.;  
 RT "A common onc gene sequence transduced by avian carcinoma virus MH2  
 RT and by murine sarcoma virus 3611.";  
 RL SCIENCE 223:813-816(1984).  
 CC -1- DISEASE: BY ITSELF THE V-MIL ONCOGENE HAS ONLY WEAK TRANSFORMING  
 CC CAPACITY BUT IT ABOLISHES THE GROWTH FACTOR REQUIREMENTS OF AVIAN  
 CC MACROPHAGES TRANSFORMED BY OTHER ONCOGENES.  
 CC -1- THIS PROTEIN IS SYNTHESIZED AS A GAG-MHT OR GAG-MIL POLYPROTEIN.  
 CC -1- SIMILARITY: WITH THE CONSERVED CATALYTIC DOMAINS OF SER/THR-  
 CC PROTEIN KINASES. BELONGS TO THE MIL/RAF SUBFAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: X00534; E235847; ALT\_INIT.  
 DR PIR: A00639; TVFVNM.  
 DR PROSITE: PS00107; PROTEIN\_KINASE\_ATP; 1.  
 DR PROSITE: PS00108; PROTEIN\_KINASE\_ST; 1.  
 DR PROSITE: PS50011; PROTEIN\_KINASE\_DOM; 1.  
 DR PFAM: PF00069; Pkinase; 1.  
 DR HSSP: P1362; IFGI.  
 KM POLYPROTEIN; TRANSFORMING PROTEIN; SERINE/THREONINE-PROTEIN KINASE;  
 KM TRANSFERASE; ONCOGENE; ATP-BINDING.  
 FT DOMAIN 82 341 PROTEIN KINASE.  
 FT NP\_BIND 88 96 ATP (BY SIMILARITY).  
 FT BINDING 108 108 ATP (BY SIMILARITY).  
 FT ACT SITE 201 201 BY SIMILARITY.  
 FT CONFLICT 211 211 G -> E (IN REF. 2).  
 SQ SEQUENCE 380 AA; 42853 MW; 1D7B2AB4 CRC32;

Query Match 41.9%; Score 62; DB 1; Length 380;  
 Best Local Similarity 30.0%; Pred. No. 4.23e+00;  
 Matches 6; Conservative 7; Mismatches 7; Indels 0; Gaps 0;  
 Db 146 TKDNLAIYQWCGSSLYKH 165  
 OY 1 ANDHLSLEAMSDNDPYLH 20

Search completed: Tue Jan 11 15:37:35 2000  
 Job time : 9 secs.

\*\*\*\*\*  
 NWSETH (TM)  
 \*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
 Copyright (c) 1993-1998 University of Edinburgh, U.K.  
 Distribution rights by Oxford Molecular Ltd

MSrch\_pp protein - protein database search, using Smith-Waterman algorithm  
 Run on: Tue Jan 11 15:47:05 2000; Maspar time 3.48 Seconds  
 253.134 Million cell updates/sec  
 No output not generated.

Title: >US-09-290-049-3  
 Description: (1-22) from US09290049.pep  
 Perfect Score: 155  
 Sequence: 1 TGARTINGOLLYFRANGVQVKG 22

Scoring table: PAM 150  
 Gap 15

Searched: 122810 seqs, 40068593 residues

Post-processing: Minimum Match 0%  
 Listing first 45 summaries

Database: p1r60  
 1:p1r1 2:p1r2 3:p1r3 4:p1r4

Statistics: Mean 29.032; Variance 42.944; scale 0.676

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
1	155	100.0	1475	2	B33135	gtfB protein precursor	4.92e-20
2	112	72.3	1592	2	A38175	glucosyltransferase p	5.30e-10
3	109	70.3	1518	2	A44811	glucosyltransferase (	2.45e-09
4	107	69.0	1290	2	JC5473	dextranucrase (EC 2.	6.75e-09
5	106	68.4	1375	2	TU0345	dextranucrase (EC 2.	1.12e-08
6	106	68.4	1599	2	S22737	glucosyltransferase (	1.12e-08
7	104	67.1	1365	2	A41483	glucosyltransferase (	3.05e-08
8	93	60.0	1431	2	A45866	dextranucrase (EC 2.	6.72e-06
9	88	56.8	1431	2	A37184	glucan-binding protei	7.21e-05
10	83	53.5	329	2	A55221	dextranase inhibitor p	7.32e-04
11	79	51.0	382	2	S55166	hypothetical protein	4.46e-03
12	77	49.7	1329	2	F71014	probable dipeptide tr	1.08e-02
13	72	46.5	126	2	S40016	doc protein - phase p	9.45e-02
14	71	45.8	425	2	H70410	UDP-N-acetylglucosam	1.44e-01
15	68	43.9	831	3	T00323	chitinase (EC 3.2.1.1	5.03e-01
16	67	43.2	245	2	F64465	hypothetical protein	7.58e-01
17	67	43.2	2710	2	A37052	toxlin A - Clostridium	7.58e-01
18	65	41.9	822	2	B54846	fibroblast growth fac	1.70e+00
19	65	41.9	1396	2	S36851	L-shaped tail fiber p	1.70e+00
20	63	40.6	240	2	H69366	conserved hypothetical	3.75e+00
21	63	40.6	301	2	S13165	astalogycoprotein re	3.75e+00
22	63	40.6	310	2	F71020	probable fructokinase	3.75e+00
23	63	40.6	336	1	WZBE8	dUTP pyrophosphatase	3.75e+00

## ALIGNMENTS

RESULT ENTRY	1	B33135	#type complete	Gal beta-1, 3galNAc-s	3.75e+00
TITLE	gtfB protein precursor - Streptococcus mutans			chlafrin-associated p	3.75e+00
ORGANISM	Streptococcus mutans			high affinity sulfate	3.75e+00
DATE	23-Oct-1990	#sequence=revision 23-Oct-1990	#text_change	inner layer protein V	3.75e+00
ACCESSIONS	B33135; A33128			L2 protein - human pa	5.53e+00
REFERENCE	A33135			ispB protein - f1ssio	5.53e+00
#authors	Shiroza, T.; Ueda, S.; Kuramitsu, H.K.			L2 protein - human pa	5.53e+00
#journal	J. Bacteriol. (1987) 169:4263-4270			cytochrome L - Clostri	5.53e+00
#cross-references	NCBI:87308013			uroporphyrin-III c-me	8.12e+00
#accession	B33135			int protein - phage 1	8.12e+00
#status	preliminary			hypothetical protein	8.12e+00
#molecule-type	DNA			hypothetical protein	8.12e+00
#residues	1-1475	#label SHI		ABC-type transport pr	8.12e+00
#cross-references	GB:M17361; NID:gl53639; PID:gl53640			cannabinoind receptor	1.19e+01
REFERENCE	A33128			cannabinoind receptor	1.19e+01
#authors	Shiroza, T.; Ueda, S.; Kuramitsu, H.K.			hypothetical protein	1.19e+01
#submission	submitted to the Protein Sequence Database, September 1990			enterotoxin A - Clost	1.19e+01
#accession	A33128			transcription factor	1.19e+01
#status	preliminary; not compared with conceptual translation			gene try protein - ra	1.19e+01
#molecule-type	DNA			regulatory protein Sp	1.19e+01
#residues	1-171,173-641,'N',643-1475	#label SH2			1.19e+01
#cross-references	1-171,173-641,'N',643-1475	#label SH2			1.19e+01
#experimental source	strain GS-5				1.19e+01
#superfamily	cp1 repeat homology				1.19e+01
FEATURE	1096-1115	#domain cpl repeat homology #label CP1\			1.19e+01
	1224-1243	#domain cpl repeat homology #label CP2\			1.19e+01
	1299-1308	#domain cpl repeat homology #label CP3\			1.19e+01
	1354-1373	#domain cpl repeat homology #label CP4\			1.19e+01
	1419-1438	#domain cpl repeat homology #label CP5			1.19e+01
SUMMARY	#length 1475	#molecular-weight 165811	#checksum 7497		1.19e+01

Query Match 100.0%; Score 155; DB 2; Length 1475;  
 Best local similarity 100.0%; Pred. No. 4.92e-20;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1300 TGARTINGOLLYFRANGVQVKG 1321  
 1 TGARTINGOLLYFRANGVQVKG 22

RESULT 2  
 ENTRY A38175 #type complete  
 TITLE glucosyltransferase precursor - Streptococcus sobrinus  
 ORGANISM Streptococcus sobrinus

DATE 28-Aug-1992 #sequence\_revision 28-Aug-1992 #text\_change 09-Sep-1997

ACCESSIONS A38175

ENTRY A38175

TITLE Abo, H.; Mtsunura, T.; Kodama, T.; Ohta, H.; Fukui, K.; Kato, K.; Kagawa, H.

ORGANISM J. Bacteriol. (1991) 173:989-996

REFERENCE Peptide sequences for sucrose splitting and glucan binding within Streptococcus sobrius glucosyltransferase (water-insoluble glucan synthetase).

#journal #title

#cross-references MUID:91123227

#accession A38175

#status Preliminary

#molecule\_type DNA

#residues 1-1592 ##label ABO

CLASSIFICATION #superfamily cpl repeat homology

FEATURE

1093-1112 #domain cpl repeat homology #label CP1\

1222-1241 #domain cpl repeat homology #label CP2\

1287-1306 #domain cpl repeat homology #label CP3\

1330-1351 #domain cpl repeat homology #label CP4\

1352-1371 #domain cpl repeat homology #label CP5\

1402-1420 #domain cpl repeat homology #label CP6\

1465-1484 #domain cpl repeat homology #label CP7\

1513-1532 #domain cpl repeat homology #label CP8\

SUMMARY #length 1592 #molecular-weight 176167 #checksum 5940

Query Match 72.3%; Score 112; DB 2; Length 1592;

Best Local Similarity 77.3%; Pred. No. 5.30e-10;

Matches 17; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Db 1298 TGAATGKQKLYFRANGOVKG 1319

QY 1 TGARTINGQLLYFRANGOVKG 22

RESULT 3

ENTRY A44811

TITLE #type complete

ORGANISM glucosyltransferase (EC 2.4.1.-) I - Streptococcus salivarius

DATE 31-Dec-1993 #sequence\_revision 31-Dec-1993 #text\_change 09-Sep-1997

ACCESSIONS A44811; S22726; S28809

REFERENCE A44811

#authors Giffard, P.M.; Simpson, C.L.; Miyard, C.P.; Jacques, N.A.

#journal J. Gen. Microbiol. (1991) 137:2577-2593

#title Molecular characterization of a cluster of at least two glucosyltransferase genes in Streptococcus salivarius ATCC 25975.

#cross-references MUID:92148377

#accession A44811

#molecule\_type DNA

#residues 1-1518 ##label GIF

GENETICS #note Sequence extracted from NCBI backbone (NCBIN:81050, NCBI:81052)

CLASSIFICATION #superfamily cpl repeat homology

KEYWORDS glycosyltransferase; hexosyltransferase

FEATURE

1307-1326 #domain cpl repeat homology #label CP4

SUMMARY #length 1518 #molecular-weight 167730 #checksum 1334

Query Match 70.3%; Score 109; DB 2; Length 1518;

Best Local Similarity 72.7%; Pred. No. 2.45e-09;

Matches 16; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Db 1383 TGAQVINGOHLYFNADSGOVKG 1404

QY 1 TGARTINGQLLYFRANGOVKG 22

RESULT 4

ENTRY JC5473

TITLE #type complete

ORGANISM dextranase (EC 2.4.1.5) - Leuconostoc mesenteroides

DATE 07-Jul-1997 #sequence\_revision 29-Aug-1997 #text\_change 29-Aug-1997

ACCESSIONS JC5473

REFERENCE JC5473

#authors Monchois, V.; Willemot, R.M.; Renaud-Simeon, M.; Croux, C.; Monnan, P.

#journal Gene (1996) 182:23-32

#title Cloning and sequencing of a gene coding for a novel dextranase from Leuconostoc mesenteroides NRRL B-1299 synthesizing only alpha (1-6) and alpha (1-3) linkages.

#cross-references MUID:97136686

#accession JC5473

#status nucleic acid sequence not shown; translation not shown

#molecule\_type DNA

#residues 1-1290 ##label MON

COMMENT This enzyme catalyzes the transfer of D-glucopyranosyl units from sucrose onto acceptor molecules.

GENETICS dsra

KEYWORDS glycosyltransferase; hexosyltransferase

FEATURE

78-870 #domain catalytic #status predicted #label CAT\

SUMMARY #length 1290 #molecular-weight 145572 #checksum 9386

Query Match 69.0%; Score 107; DB 2; Length 1290;

Best Local Similarity 63.6%; Pred. No. 6.75e-09;

Matches 14; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Db 1181 TGDRIIRGQNLRYNONGIOMKG 1202

QY 1 TGARTINGQLLYFRANGOVKG 22

RESULT 5

ENTRY JT0345

TITLE #type complete

ORGANISM dextranase (EC 2.4.1.5) precursor - Streptococcus mutans (strain GS-5)

DATE 31-Mar-1992 #sequence\_revision 31-Mar-1992 #text\_change 22-Nov-1996

ACCESSIONS JT0345; C33135

REFERENCE JT0345

#authors Ueda, S.; Shiroya, T.; Kuramitsu, H.K.

#journal Gene (1988) 69:101-109

#title Sequence analysis of the glfC gene from Streptococcus mutans GS-5

#cross-references MUID:89137980

#accession JT0345

#molecule\_type DNA

#residues 1-1375 ##label UED

GENETICS #experimental source GS-5

REFERENCE A33135

#authors Shiroya, T.; Ueda, S.; Kuramitsu, H.K.

#journal J. Bacteriol. (1987) 169:4263-4270

#title Sequence analysis of the glfB gene from Streptococcus mutans.

#cross-references MUID:87308013

#accession C33135

#status Preliminary

#molecule\_type DNA

#residues 1-349 ##label SHI

GENETICS #cross-references GB:M17361

FUNCTION #gene glfC

#description catalyzes the synthesis of both water-soluble and



```

#title      Sequence analysis of the gene for the glucan-binding protein
#cross-references MUID:90170123
#accession  A37184
#status     preliminary
#molecule_type DNA
#residues   1563 #label BAN
#cross-references GB:M50945; MID:g153637; PID:g153638
CLASSIFICATION #superfamily cpl repeat homology
FEATURE
169-188      #domain cpl repeat homology #label CP1\
264-283      #domain cpl repeat homology #label CP2\
349-368      #domain cpl repeat homology #label CP3\
504-523      #domain cpl repeat homology #label CP4\
525-548      #domain cpl repeat homology #label CP5
SUMMARY
length 563 #molecular-weight 62932 #checksum 2435

Query Match      56.88; Score 88; DB 2; Length 563;
st. Local Similarity 68.48; Pred. No. 7.21e-05;
cches 13; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

Db 278 RTIGKKRYFDNGQYKG 296
||| | | | | | | | |
4 RTINGQLLYFRANGQYKG 22

RESULT 10
ENTRY      A55221
TITLE      #type complete
ORGANISM    dextranase inhibitor precursor - Streptococcus sobrinus
            (strain UAB66, serotype g)
DATE        23-Mar-1995 #sequence_revision 23-Mar-1995 #text_change
17-Mar-1999
ACCESSIONS  A55221
REFERENCE    Sun, J.W.; Wanda, S.Y.; Camilli, A.; Curtiss III, R.
            J. Bacteriol. (1994) 176:7213-7222
            Cloning and DNA sequencing of the dextranase inhibitor gene
            (del) from Streptococcus sobrinus.
            #cross-references MUID:95050304
            #accession A55221
            #status     preliminary
            #molecule_type DNA
            #residues   1-329 #label SUN
            #cross-references GB:L34406
CLASSIFICATION #superfamily cpl repeat homology
FEATURE
173-192      #domain cpl repeat homology #label CPl
            #length 329 #molecular-weight 36077 #checksum 8218

Query Match      53.58; Score 83; DB 2; Length 329;
Best Local Similarity 68.28; Pred. No. 7.32e-04;
Matches 15; Conservative 2; Mismatches 4; Indels 1; Gaps 1;

Db 185 GLQINGQTLFDPNTRQYKG 206
| : | | | | | | | | | | | | |
2 GARTINGQLLYFRAN-GVQYKG 22

RESULT 11
ENTRY      S55166
TITLE      #type complete
            hypothetical protein YOL147c - yeast (Saccharomyces
            cerevisiae)
ALTERNATE_NAMES #formal_name Saccharomyces cerevisiae
ORGANISM        08-Jul-1995 #sequence_revision 01-Sep-1995 #text_change
DATE            05-Jun-1998
ACCESSIONS      S55166; S56929; S71650
REFERENCE        Katsoulou, C.; Tzermia, M.; Alexandraki, D.
            submitted to the EMBL Data Library, May 1995
            #authors
            #description The complete sequence of a 40.7 kb segment located on the
            left arm of yeast chromosome X identified 13 known genes

```

```

#accession  S55166
#molecule_type DNA
#residues   1-382 #label KAT
#cross-references EMBL:X87371; MID:9854542; PID:9854550
REFERENCE      S56912
#authors       Katsoulou, C.; Tzermia, M.; Alexandraki, D.
#submissions   submitted to the Protein Sequence Database, September 1995
#accession     S56929
#molecule_type DNA
#residues     1-382 #label KAM
#cross-references EMBL:Z49422; MID:g1015560; PID:g1015561; MIPS:YTL147C
REFERENCE      S71643
#authors       Katsoulou, C.; Tzermia, M.; Tavernarakis, N.; Alexandraki, D.
#journal        Yeast (1996) 12:787-797
#title          Sequence analysis of a 40.7 kb segment from the left arm of
            yeast chromosome X reveals 14 known genes and 13 new open
            reading frames including homologues of genes clustered on
            the right arm of chromosome XI.

#accession     S71650
#status        nucleic acid sequence not shown; translation not shown
#molecule_type DNA
#residues     1-382 #label KAF
#cross-references EMBL:X87371; MID:9854542; PID:9854550
#note          the nucleotide sequence was submitted to the EMBL Data
            Library, May 1995

GENETICS
#map_position 101.
SUMMARY        #length 382 #molecular-weight 44862 #checksum 8833

Query Match      51.08; Score 79; DB 2; Length 382;
Best Local Similarity 38.98; Pred. No. 4.46e-03;
Matches 7; Conservative 7; Mismatches 4; Indels 0; Gaps 0;

Db 170 RYVVISILRYRTDQIR 187
| : | | | | | | | | |
4 RTINGQLLYFRANGQYKG 21

RESULT 12
ENTRY      F71014
TITLE      #type complete
ORGANISM    probable dipeptide transport ATP-binding protein dppd -
            Pyrococcus horikoshii
DATE        14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change
21-Nov-1998
ACCESSIONS  F71014
REFERENCE    Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Halkawa, Y.;
            Hino, Y.; Yamamoto, S.; Sekine, M.; Baba, S.; Kosugi, H.;
            Hosoyama, A.; Nagai, Y.; Sakai, M.; Ogura, K.; Otsuka, R.;
            Nakazawa, H.; Takamiya, M.; Ohnuki, Y.; Funahashi, T.;
            Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi,
            A.; Aoki, K.; Yoshizawa, T.; Nakamura, Y.; Robb, F.T.;
            Horikoshi, K.; Masuchi, Y.; Shizuya, H.; Kikuchi, H.
            DNA Res. (1998) 5:55-76
            Complete sequence and gene organization of the genome of a
            hyper-thermophilic archaeobacterium, Pyrococcus horikoshii
            OT3.
            #journal
            #title          preliminary; nucleic acid sequence not shown;
            translation not shown
            #accession     F71014
            #cross-references MUID:98344137
            #status        preliminary; nucleic acid sequence not shown;
            translation not shown
            #molecule_type DNA
            #residues     1-329 #label KAM
            #cross-references GB:AP000006; MID:g3236133; PID:d1031461; PID:g3257835
            #experimental_source strain OT3
            #note          this accession replaces an interim accession for a
            sequence replaced by Genbank

GENETICS
#note          PH1412
FUNCTION

```

#description probably responsible for energy-coupling to the transport system

CLASSIFICATION #superfamily inner membrane protein malK; ATP-binding cassette homology

KEYWORDS ATP; binding protein-dependent transport system; nucleotide binding; oligopeptide transport; P-loop

FEATURE

28-241 #domain ATP-binding cassette homology #label ABC\

45-53 #region nucleotide-binding motif A (P-loop)\

185-189 #region nucleotide-binding motif B

SUMMARY #length 329 #molecular-weight 36504 #checksum 6349

Query Match 49.78; Score 77; DB 2; Length 329;

Best Local Similarity 44.48; Pred. No. 1.08e-02;

Matches 8; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

Db 66 GRVGGVLYRREDGSVY 83

QY 3 ARTNGQLYFRANGVOV 20

ENTRY 13

TITLE S40016 #type complete

ORGANISM doc protein - phage P1

DATE 25-Dec-1994 #sequence-revision 17-May-1996 #text-change 29-Jan-1999

ACCESSIONS S40016; S38555

REFERENCE S40015

#authors Lehnher, H.; Maguin, E.; Jafri, S.; Yarmolinsky, M.B.

#journal J. Mol. Biol. (1993) 233:414-428

#title Plasmid addiction genes of bacteriophage P1: doc, which causes cell death on curing of prophage, and phd, which prevents host death when prophage is retained.

#cross-references M01D:94016561

#accession S40016

#molecule\_type DNA

#residues 1-126 #label LEH

REFERENCE S38553

#cross-references GB:M95666; NID:g463276; PID:g463277

#authors Schmidt, C.; Lehnher, H.; Guldolin, A.; Aber, W.

#submission Submitted to the EMBL Data Library, November 1992

#description Additional late promoter sequences of bacteriophage P1.

#accession S38555

#molecule\_type DNA

#residues 26-126 #label SCH

#cross-references EMBL:M95666

SUMMARY #length 126 #molecular-weight 13588 #checksum 6874

Query Match 46.58; Score 72; DB 2; Length 126;

Best Local Similarity 56.38; Pred. No. 9.45e-02;

Matches 9; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

Db 76 ALNSALLFRNGVOV 91

QY 5 TINGQLYFRANGVOV 20

RESULT 14

ENTRY H70410 #type complete

TITLE UDP-N-acetylglucosamine 1-carboxyvinyltransferase - Aquifex aeolicus

ORGANISM #formal\_name Aquifex aeolicus

DATE 08-May-1998 #sequence-revision 08-May-1998 #text-change 13-Sep-1998

ACCESSIONS H70410

REFERENCE A70300

#authors Decker, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; Overbeek, R.; Sneed, M.A.; Keller, M.; Aulay, M.; Huber, R.; Feldman, R.A.; Short, J.M.; Olson, G.J.; Swanson, R.V.

#journal Nature (1998) 392:353-358

#title The complete genome of the hyperthermophilic bacterium

Aquifex aeolicus.

#cross-references M01D:98196666

#accession H70410

#status preliminary; nucleic acid sequence not shown; translation not shown

#molecule\_type DNA

#residues 1-425 #label AOF

#cross-references GB:AE000732; NID:g2983704; PID:g2983705; GB:AE000657

#experimental\_source strain VF5

GENETICS

#gene mura

CLASSIFICATION #superfamily UDP-N-acetylglucosamine 1-carboxyvinyltransferase MurZ

SUMMARY #length 425 #molecular-weight 47259 #checksum 7026

Query Match 45.88; Score 71; DB 2; Length 425;

Best Local Similarity 42.18; Pred. No. 1.44e-01;

Matches 8; Conservative 5; Mismatches 6; Indels 0; Gaps 0;

Db 127 GARPIDHLEKFEKAGADV 145

QY 2 GARTINGQLYFRANGVOV 20

RESULT 15

ENTRY T00323 #type complete

TITLE chitinase (EC 3.2.1.14) B - Clostridium paraputrificum

ORGANISM #formal\_name Clostridium paraputrificum

DATE 29-Jan-1999 #sequence-revision 29-Jan-1999 #text-change 29-Jan-1999

ACCESSIONS T00323

REFERENCE T00323

#authors Ohmiva, K.; Sakka, K.; Kimura, T.; Karita, S.; Morimoto, K.

#journal J. Bacteriol. (1997) 179:7306-7314

#title Cloning, sequencing, and expression of the gene encoding Clostridium paraputrificum chitinase ChA and analysis of the functions of novel cadherin-like domains and a chitin-binding domain.

#accession T00323

#status preliminary

#residues 1-831 #label OHM

#cross-references EMBL:AB001874; NID:d1175645; PID:d1024701

SUMMARY #length 831 #molecular-weight 90021 #checksum 7845

Query Match 43.98; Score 68; DB 3; Length 831;

Best Local Similarity 36.48; Pred. No. 5.03e-01;

Matches 8; Conservative 9; Mismatches 5; Indels 0; Gaps 0;

Db 525 TAARTVKNMKVYIKTNGVAITG 546

QY 1 TGARTINGQLYFRANGVOVG 22

Search completed: Tue Jan 11 15:47:33 2000

Job time : 28 secs.







```

ALTERNATE_NAMES  sucrose 6-glucosyltransferase
ORGANISM          #formal_name Streptococcus mutans
DATE              31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change
                22-Nov-1996
ACCESSIONS       JT0345; C33135
REFERENCE        JT0345
#authors         Ueda, S.; Shiroza, T.; Kuramitsu, H.K.
#journal         Gene (1988) 69:101-109
#title           Sequence analysis of the gtfC gene from Streptococcus mutans
#               GS-5.
#cross-references MUID:89137980
#accession       JT0345
#molecule_type DNA
#residues        1-1375 ##label UED
#experimental_source GS-5
REFERENCE        A33135
#authors         Shiroza, T.; Ueda, S.; Kuramitsu, H.K.
#journal         J. Bacteriol. (1987) 169:4263-4270
#title           Sequence analysis of the gtfB gene from Streptococcus mutans.
#cross-references MUID:87308013
#accession       C33135
#status          preliminary
#molecule_type DNA
#residues        1-1349 ##label SHI
#cross-references GB:M17361
GENETICS
#gene            gtfC
#function         catalyzes the synthesis of both water-soluble and
                  water-insoluble glucans from glucose.
CLASSIFICATION  #superfamily cpl repeat homology
KEYWORDS         duplication; glycosyltransferase; hexosyltransferase;
                  transferase
FEATURE
#               #domain signal sequence #status predicted #label SIG\
1-34            #product glucosyltransferase #status predicted #label
35-1375         MAT\
1126-1135       #domain cpl repeat homology #label CP1\
1253-1272       #domain cpl repeat homology #label CP2\
1318-1337       #domain cpl repeat homology #label CP3
SUMMARY         #length 1375 #molecular-weight 153021 #checksum 7015
Query Match     91.9%; Score 156; DB 2; Length 1375;
Best Local Similarity 95.2%; Pred. No. 3,136-14;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Db 507 ANDHLSTLEAMSYNDPYLHD 527
1 ANDHLSTLEAMSYNDPYLHD 21
RESULT 3
ENTRY      A38175 #type complete
TITLE      glucosyltransferase precursor - Streptococcus sobrinus
ORGANISM   #formal_name Streptococcus sobrinus
DATE       28-Aug-1992 #sequence_revision 28-Aug-1992 #text_change
09-Sep-1997
ACCESSIONS A38175
REFERENCE   A38175
#authors   Abo, H.; Mtsunura, T.; Kodama, T.; Ohta, H.; Fukui, K.; Kato,
            K.; Kagawa, H.
#journal   J. Bacteriol. (1991) 173:989-996
#title     Peptide sequences for sucrose splitting and glucan binding
            within Streptococcus sobrinus glucosyltransferase
            (water-insoluble glucan synthetase).
#cross-references MUID:91123227
#accession  A38175
#status     preliminary
#molecule_type DNA
#residues   1-1592 ##label ABO
#cross-references GB:D90213; NID:9217032; PID:d1014946; PID:9217033
CLASSIFICATION #superfamily cpl repeat homology
FEATURE

```

```

1093-1112       #domain cpl repeat homology #label CP1\
1222-1241       #domain cpl repeat homology #label CP2\
1287-1306       #domain cpl repeat homology #label CP3\
1330-1351       #domain cpl repeat homology #label CP4\
1352-1371       #domain cpl repeat homology #label CP5\
1402-1420       #domain cpl repeat homology #label CP6\
1465-1484       #domain cpl repeat homology #label CP7\
1513-1532       #domain cpl repeat homology #label CP8
SUMMARY         #length 1592 #molecular-weight 176167 #checksum 5940
Query Match     88.5%; Score 131; DB 2; Length 1592;
Best Local Similarity 85.7%; Pred. No. 3,816-13;
Matches 18; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
Db 477 ANNEVSIEMASNDPYLHD 497
1 ANDHLSTLEAMSYNDPYLHD 21
RESULT 4
ENTRY      A41483 #type complete
TITLE      glucosyltransferase (EC 2.4.1.-) gtfS precursor -
ORGANISM   Streptococcus sobrinus
DATE       30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change
09-Sep-1997
ACCESSIONS A41483
REFERENCE   A41483
#authors   Gilmore, K.S.; Russell, R.R.B.; Ferretti, J.J.
#journal   Infect. Immun. (1990) 58:2452-2458
#title     Analysis of the Streptococcus downei gtfS gene, which
            specifies a glucosyltransferase that synthesizes soluble
            glucans.
#cross-references MUID:90316665
#accession  A41483
#molecule_type DNA
#residues   1-1365 ##label GIL
#cross-references GB:M30943; NID:9153652; PID:9153653
GENETICS
#gene       gtfS
#function   #superfamily cpl repeat homology
KEYWORDS    glycosyltransferase; hexosyltransferase
SUMMARY     #length 1365 #molecular-weight 151590 #checksum 823
Query Match     68.9%; Score 102; DB 2; Length 1365;
Best Local Similarity 71.4%; Pred. No. 4,406-07;
Matches 15; Conservative 1; Mismatches 5; Indels 0; Gaps 0;
Db 467 AIDHLSTLEAMSGNDPYLHD 487
1 ANDHLSTLEAMSYNDPYLHD 21
RESULT 5
ENTRY      A45866 #type complete
TITLE      dextransucrase (EC 2.4.1.5) precursor - Streptococcus mutans
ORGANISM   #formal_name Streptococcus mutans
DATE       30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change
07-Jul-1995
ACCESSIONS A45866
REFERENCE   A45866
#authors   Honda, O.; Kato, C.; Kuramitsu, H.K.
#journal   J. Gen. Microbiol. (1990) 136:2099-2105
#title     Nucleotide sequence of the streptococcus mutans gtdF gene
            encoding the glucosyltransferase-S enzyme.
#accession  A45866
#status     preliminary
#molecule_type DNA
#residues   1-1431 ##label HON
#cross-references GB:M29296
CLASSIFICATION #superfamily cpl repeat homology
KEYWORDS      glycosyltransferase; hexosyltransferase
FEATURE

```

181-201 #domain cpl repeat homology #label CP1  
1127-1146 #domain cpl repeat homology #label CP2  
1192-1211 #domain cpl repeat homology #label CP3  
1257-1276 #domain cpl repeat homology #label CP4  
1277-1297 #domain cpl repeat homology #label CP5  
1321-1340 #domain cpl repeat homology #label CP6  
1341-1361 #domain cpl repeat homology #label CP7  
1385-1404 #domain cpl repeat homology #label CP7  
SUMMARY #length 1431 #molecular-weight 159844 #checksum 4333

Query Match 68.2% Score 101; DB 2; Length 1431;  
Best Local Similarity 77.8%; Pred. No. 6,99e-07;  
Matches 14; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

DB 495 AINHLSEAWSDNDPXY 512  
1:|||||  
OY 1 ANDHLSLEAWSDNDPXY 18

RESULT 6  
ENTRY S22737 #type complete  
TITLE glucosyltransferase (EC 2.4.1.-) S - Streptococcus salivarius  
ORGANISM #formal\_name Streptococcus salivarius  
DATE 16-Sep-1992 #sequence\_revision 16-Sep-1992 #text\_change  
09-Sep-1997  
ACCESSIONS S22737; S28810; B44811; S22727  
REFERENCE S22726  
#authors Jacques, N.  
#submission submitted to the EMBL Data Library, March 1992  
#accession S22737  
#molecule\_type DNA  
#residues 1-1599 #label JAC  
#cross-references EMBL:Z11872; NID:g47530; PID:g47531  
#experimental\_source ATCC 25975

REFERENCE A44811  
#authors Giffard, P.M.; Simpson, C.L.; Milward, C.P.; Jacques, N.A.  
#journal J. Gen. Microbiol. (1991) 137:2577-2593  
#title Molecular characterization of a cluster of at least two glucosyltransferase genes in Streptococcus salivarius ATCC 25975.

#cross-references MVID:92148377  
#accession S28810  
#molecule\_type DNA  
#residues 1-51 #label GIF  
#cross-references EMBL:Z11873

GENETICS  
#gene gtfK  
CLASSIFICATION #superfamily cpl repeat homology  
KEYWORDS glucosyltransferase; hexosyltransferase  
SUMMARY #length 1599 #molecular-weight 176480 #checksum 3135

Query Match 52.7% Score 78; DB 2; Length 1599;  
Best Local Similarity 55.6%; Pred. No. 1.70e-02;  
Matches 10; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

DB 494 NISLEAWSHNDPYVNE 511  
:|||||  
OY 4 HLSLEAWSDNDPXYLHD 21

RESULT 7  
ENTRY A44811 #type complete  
TITLE glucosyltransferase (EC 2.4.1.-) I - Streptococcus salivarius  
ORGANISM #formal\_name Streptococcus salivarius  
DATE 31-Dec-1993 #sequence\_revision 31-Dec-1993 #text\_change  
09-Sep-1997  
ACCESSIONS A44811; S22726; S28809  
REFERENCE A44811  
#authors Giffard, P.M.; Simpson, C.L.; Milward, C.P.; Jacques, N.A.  
#journal J. Gen. Microbiol. (1991) 137:2577-2593  
#title Molecular characterization of a cluster of at least two glucosyltransferase genes in Streptococcus salivarius ATCC 25975.

#cross-references MVID:92148377  
#accession A44811  
#molecule\_type DNA  
#residues 1-1518 #label GIF  
#cross-references EMBL:Z11873; NID:g47526; PID:g47527  
#note Sequence extracted from NCBI Backbone (NCBIN:81050, NCBI:81052)

GENETICS  
#gene gtfJ  
CLASSIFICATION #superfamily cpl repeat homology  
KEYWORDS glucosyltransferase; hexosyltransferase  
FEATURE 1307-1326  
SUMMARY #length 1518 #molecular-weight 167730 #checksum 1334

Query Match 50.7% Score 75; DB 2; Length 1518;  
Best Local Similarity 66.7%; Pred. No. 5.82e-02;  
Matches 10; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

DB 504 HISVLEAWSLDNHXY 518  
1:|||||  
OY 4 HLSLEAWSDNDPXY 18

RESULT 8  
ENTRY B48445 #type complete  
TITLE glyceraldehyde-3-phosphate dehydrogenase (EC 1.2.1.12) -  
Leishmania mexicana  
ORGANISM #formal\_name Leishmania mexicana  
DATE 31-Dec-1993 #sequence\_revision 31-Dec-1993 #text\_change  
08-Sep-1997  
ACCESSIONS B48445; S25142  
REFERENCE B48445  
#authors Hannevert, V.; Blauw, M.; Kohl, L.; Allert, S.; Oppertdoes, F.R.; Michels, P.A.M.;  
#journal Mol. Biochem. Parasitol. (1992) 55:115-126  
#title Molecular analysis of the cytosolic and glycosomal glyceraldehyde-3-phosphate dehydrogenase in Leishmania mexicana.

#cross-references MVID:93063042  
#accession B48445  
#status preliminary  
#molecule\_type DNA  
#residues 1-331 #label HAN  
#cross-references EMBL:X65220; NID:g9552; PID:g9553

CLASSIFICATION #superfamily glyceraldehyde-3-phosphate dehydrogenase  
KEYWORDS oxidoreductase  
SUMMARY #length 331 #molecular-weight 35643 #checksum 147

Query Match 48.6% Score 72; DB 2; Length 331;  
Best Local Similarity 52.6%; Pred. No. 1.94e-01;  
Matches 10; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

DB 301 NDHFVRLVSWYDNEGYSH 319  
|||:|:|:|:|  
OY 2 NDHLSLEAWSDNDPXYLH 20

RESULT 9  
ENTRY OQB640 #type complete  
TITLE BGLF2 protein - human herpesvirus 4 (strain B95-8)  
ORGANISM #formal\_name human herpesvirus 4, Epstein-Barr virus  
DATE 25-Feb-1985 #sequence\_revision 25-Feb-1985 #text\_change  
22-Jan-1999  
ACCESSIONS C43044; J01381; A03784; A03794; S33036  
REFERENCE A93065  
#authors Bankier, A.T.; Deininger, P.L.; Farrell, P.J.; Barrell, B.G.;  
#journal Mol. Biol. Med. (1983) 1:21-45  
#title Sequence analysis of the 17,166 bp EcoRI fragment C of B95-8 Epstein-Barr virus.

#cross-references MVID:85035713  
#accession C43044  
#molecule\_type DNA

##residues 1-336 ##label BAN  
##cross-references EMBL:V01555; NID:G59074; PID:G1334895  
REFERENCE A03794  
#authors Baer, R.; Bankier, A.T.; Biggin, M.D.; Deininger, P.L.;  
Farrell, P.J.; Gibson, T.J.; Hatfull, G.; Hudson, G.S.;  
Satchwell, S.C.; Seglin, C.; Tuffnell, P.S.; Barrett, B.G.  
#journal Nature (1984) 310:207-211  
#title DNA sequence and expression of the B95-8 Epstein-Barr virus  
genome.  
#cross-references MUID:84270667  
#contents annotation: protein coding region  
#note neither amino acid nor nucleotide sequence is given  
REFERENCE J01381  
#authors Chen, M.R.; Hsu, T.Y.; Lin, S.W.; Chen, J.Y.; Yang, C.S.  
#journal J. Gen. Virol. (1991) 72:3047-3055  
#title Cloning and characterization of cDNA clones corresponding to  
transcripts from the BamHI G region of the Epstein-Barr  
virus genome and expression of BGLF2.  
#cross-references MUID:92113548  
accession J01381  
##molecule\_type mRNA  
##residues 1-336 ##label CHE  
##cross-references GB:S77132; NID:G243314; PID:G243315  
CLASSIFICATION #superfamily Epstein-Barr virus Gg17 protein  
SUMMARY #length 336 #molecular-weight 36888 #checksum 8731  
Query Match 48.6%; Score 72; DB 1; Length 336;  
Best Local Similarity 41.2%; Pred. No. 1.94e-01;  
Matches 7; Conservative 7; Mismatches 3; Indels 0; Gaps 0;  
Db 216 AGAIVNLRGWTEDDSP 232  
QY 1 ANDHLEAMSDNDTP 17  
RESULT 10  
ENTRY E71557 #type complete  
TITLE hypothetical protein CT085 - Chlamydia trachomatis (serotype  
D, strain UW3/Cx)  
ORGANISM #formal\_name Chlamydia trachomatis  
DATE 13-Sep-1998 #sequence\_revision 13-Sep-1998 #text\_change  
21-Nov-1998  
ACCESSIONS E71557  
REFERENCE A71570  
#authors Stephens, R.S.; Kalman, S.; Lammel, C.J.; Fan, J.; Marathe,  
R.; Aravind, L.; Mitchell, W.P.; Olinger, L.; Tatusov,  
R.L.; Zhao, Q.; Koonin, E.V.; Davis, R.W.  
#journal Science (1998) 282:754-759  
#title Genome sequence of an obligate intracellular pathogen of  
humans: Chlamydia trachomatis.  
#cross-references MUID:9900809  
accession E71557  
##status preliminary  
##molecule\_type DNA  
##residues 1-579 ##label ARN  
##cross-references GB:AE001283; GB:AE001273; NID:G3328480; PID:G3328481  
#experimental\_source serotype D, strain UW-3/Cx  
GENETICS CT085  
SUMMARY #length 579 #molecular-weight 65354 #checksum 8323  
Query Match 48.6%; Score 72; DB 2; Length 579;  
Best Local Similarity 31.3%; Pred. No. 1.94e-01;  
Matches 5; Conservative 8; Mismatches 3; Indels 0; Gaps 0;  
Db 132 QLPMLTSMPEDEGPF 147  
QY 4 HLESLAMSDNDTPYL 19  
RESULT 11  
ENTRY H70538 #type complete  
TITLE Probable ppdk protein - Mycobacterium tuberculosis (strain

ORGANISM H37RV)  
#formal\_name Mycobacterium tuberculosis  
DATE 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change  
17-Jul-1998  
ACCESSIONS H70538  
REFERENCE A70500  
#authors Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher,  
C.; Harris, D.; Gordon, S.V.; Eigmeier, K.; Gas, S.; Barry  
III, C.E.; Tekala, F.; Badcock, K.; Basham, D.; Brown, D.;  
Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.;  
Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.;  
Hornsby, T.; Jagels, K.; Krogh, A.; McLean, J.; Moule, S.;  
Murphy, L.; Oliver, S.; Osborne, J.; Quail, M.A.;  
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.;  
Skellton, S.; Squares, S.; Squares, R.; Sulston, J.E.;  
Taylor, K.; Whitehead, S.; Barrett, B.G.  
#journal Nature (1998) 393:537-544  
#title Deciphering the biology of Mycobacterium tuberculosis from  
the complete genome sequence.  
#cross-references MUID:98295987  
accession H70538  
##status preliminary: nucleic acid sequence not shown;  
translation not shown  
##molecule\_type DNA  
##residues 1-490 ##label COL  
##cross-references GB:G95585; GB:AL123456; NID:G3261787; PID:G317147;  
PID:G2117221  
#experimental\_source strain H37RV  
GENETICS  
#gene ppdk  
SUMMARY #length 490 #molecular-weight 52004 #checksum 20  
Query Match 45.3%; Score 67; DB 2; Length 490;  
Best Local Similarity 57.1%; Pred. No. 1.36e+00;  
Matches 8; Conservative 4; Mismatches 2; Indels 0; Gaps 0;  
Db 467 LSAWSSDPPEKE 480  
QY 8 LEAMSDNDTPYLHD 21  
RESULT 12  
ENTRY JC4086 #type complete  
TITLE fatty-acid synthase (Ec 2.3.1.85) alpha chain - yeast  
(Candida albicans)  
ORGANISM #formal\_name Candida albicans  
DATE 12-Jul-1995 #sequence\_revision 03-Aug-1995 #text\_change  
05-Dec-1998  
ACCESSIONS JC4086  
REFERENCE JC4086  
#authors Southard, S.B.; Cihlar, R.L.  
#journal Gene (1995) 156:133-138  
#title Analysis and expression of the Candida albicans FAS2 gene.  
#cross-references MUID:95255657  
accession JC4086  
##molecule\_type DNA  
##residues 1-1885 ##label SOU  
##cross-references GB:L29063; NID:G456442; PID:G456443  
GENETICS  
#gene fas2  
#map\_position 3  
CLASSIFICATION #superfamily yeast fatty-acid synthase  
KEYWORDS acyltransferase; fatty acid biosynthesis; phosphopantetheine;  
phosphoprotein  
FEATURE 1299-1304  
181 #region cerulenin binding #status predicted  
#binding\_site phosphopantetheine (Ser) (covalent)  
#status predicted  
SUMMARY #length 1885 #molecular-weight 207588 #checksum 7273  
Query Match 44.6%; Score 66; DB 2; Length 1885;  
Best Local Similarity 25.0%; Pred. No. 1.98e+00;  
Matches 5; Conservative 8; Mismatches 7; Indels 0; Gaps 0;



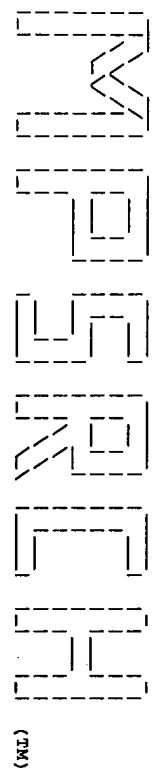
Thu Jan 13 09:13:33 2000

US-09-290-049-1.rpr

Page 6

```
Matches 7; Conservative 7; Mismatches 5; Indels 0; Gaps 0;
Db 128 TEENLEFLEARTIDRAYI 146
:::|:||||| |:|:
QY 1 ANDHLSLEAMSDNDTPYL 19
```

Search completed: Tue Jan 11 15:37:09 2000  
Job time : 14 secs.



Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MSrch\_PP protein - protein database search, using Smith-Waterman algorithm  
Run on: Tue Jan 11 15:46:28 2000; Maspar time 3.84 Seconds  
121.776 Million cell updates/sec  
ar output not generated.

Title: >US-09-290-049-3  
Description: (1-22) from US09290049.pep  
Perfect Score: 155  
Sequence: 1 TGARTINGQLYFRANGVQVKG 22

Scoring table: PAM 150  
Gap 15

Searched: 170751 seqs, 21266608 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: a-geneseq35  
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7  
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13  
14:part14 15:part15 16:part16 17:part17 18:part18  
19:part19 20:part20 21:part21 22:part22 23:part23  
24:part24 25:part25 26:part26 27:part27 28:part28  
29:part29 30:part30 31:part31 32:part32 33:part33  
34:part34 35:part35 36:part36 37:part37 38:part38  
39:part39

Statistics: Mean 20.775; Variance 64.431; scale 0.322

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
1	114	73.5	22	25	W34159	GTP antigenic peptide	4.41e-05
2	114	73.5	22	8	R43696	GTP subsequence.	4.41e-05
3	114	73.5	24	25	W34164	GTP antigenic peptide	4.41e-05
4	112	72.3	1592	6	R32925	Glucosyltransferase	7.62e-05
5	88	56.8	1577	16	R91047	Alpha-D-glucosyltrans	4.64e-02
6	67	43.2	811	17	R95014	C. difficile toxin A	9.21e+00
7	67	43.2	812	17	R95017	C. difficile toxin A	9.21e+00
8	67	43.2	2710	17	R95016	C. difficile toxin A	9.21e+00
9	67	43.2	2710	35	W68387	Clostridium difficile	9.21e+00
10	63	40.6	404	16	R78621	Chicken galNAc-4-epi	3.04e+01
11	62	40.0	349	19	W06416	Phosphotriesterase-re	3.04e+01
12	61	39.4	207	34	W38549	S. pneumoniae peptide	3.85e+01
13	61	39.4	514	39	W83359	Streptococcus pneumo	3.85e+01
14	60	38.7	116	19	W03642	Human cannabinoid GPR	4.87e+01
15	60	38.7	472	3	R14196	Human cannabinoid rec	4.87e+01
16	60	38.7	473	3	R14195	Rat cannabinoid recep	4.87e+01

17	59	38.1	91	17	R95015	C. difficile toxin A	6.14e+01
18	59	38.1	233	13	R67918	(1-3)-beta-D-glucan s	6.14e+01
19	59	38.1	453	10	R51083	fabg gene encoding bi	6.14e+01
20	59	38.1	453	20	W05208	Synechococcus biofin	6.14e+01
21	59	38.1	453	35	W70403	Synechococcus biofin	6.14e+01
22	59	38.1	673	13	R67915	(1-3)-beta-D-glucan s	6.14e+01
23	59	38.1	998	19	W03421	Mouse developmental k	6.14e+01
24	59	38.1	988	15	R85092	EPH-like receptor pro	6.14e+01
25	58	37.4	1301	6	R28304	Acid alpha-amylose en	7.74e+01
26	57	36.8	488	5	R24863	Sequence of pre-pro s	9.74e+01
27	57	36.8	492	5	R24863	Sequence of pre-pro s	9.74e+01
28	57	36.8	501	15	R89278	Cm-cellulase.	9.74e+01
29	57	36.8	501	3	R14372	Tomato endo-1,4-B-glu	9.74e+01
30	57	36.8	572	8	R40843	Bilirubin oxidase.	9.74e+01
31	56	36.1	15	9	R45658	Human progesterone re	1.22e+02
32	56	36.1	336	6	R34293	C-terminally truncated	1.22e+02
33	56	36.1	336	6	R34293	C-terminally truncated	1.22e+02
34	56	36.1	432	6	R34294	HPL(-) mutant with GP	1.22e+02
35	56	36.1	448	9	R49873	Bacillus sphaericus 5	1.22e+02
36	56	36.1	450	6	R30738	Human pancreatic lipa	1.22e+02
37	56	36.1	466	22	W20586	H. pylori cytoplasmic	1.22e+02
38	55	35.5	265	28	W46518	Amino acid sequence o	1.53e+02
39	55	35.5	449	12	R66742	Biotin-carboxylase.	1.53e+02
40	55	35.5	485	3	R13513	P.denitrificans COB Q	1.53e+02
41	55	35.5	503	3	P70219	Sequence of serrapept	1.53e+02
42	55	35.5	1139	37	W76425	Human JAK2 protein se	1.53e+02
43	55	35.5	1657	24	W18822	Human ICGAP1.	1.92e+02
44	54	34.8	667	28	W44701	Chimeric Cre-LDB-GR(I	1.92e+02
45	54	34.8	2366	17	R95011	C. difficile toxin B.	1.92e+02

ALIGNMENTS

RESULT 1  
ID W34159 standard; peptide: 22 AA.  
AC W34159; 18-FEB-1998 (first entry)  
DT 18-FEB-1998 (first entry)  
DE GTP antigenic peptide #4.  
KW Glucosyltransferase; GTP: enzyme; immunogen; catalytic domain; vaccine;  
KW surface domain; glucan-binding domain; mutans streptococcal strain;  
KW immune response; glucan-binding activity; dental caries prevention.  
OS Streptococcus mutans.  
PN US5686075-A.  
PD 11-NOV-1997.  
PF 01-MAY-1992; 877295.  
PR 30-APR-1993; US-057162.  
PR 01-MAY-1992; US-877295.  
PA (FORS-) FORSYTH DENTAL INFIRMARY FOR CHILDREN.  
PI Smith DJ, Taubman MA;  
DR WPI: 97-558089/51.  
PT Immunogenic compositions containing streptococcal  
PT glucosyl:transferase peptide(s) - used for provoking immune response  
PT to streptococcal glucosyl:transferase for preventing dental caries  
PT Claim 2: Column 13; 11pp; English.  
CC W34156-W34160 represent immunogenic fragments of the Streptomyces mutans  
CC glucosyltransferase (GTF) enzyme. W34157 and W34158 are from the  
CC catalytic domain of GTF, while W34160 is from the GTF surface domain.  
CC W34156 and W34159 are used in the glucan-binding domain of GTF. These  
CC sequences can all be used in the immunogenic composition of the  
CC invention. The composition of the invention can alternatively comprise  
CC one or more of these sequences linked to a lysine core matrix (see  
CC W34161-W34165). A composition comprising one of these sequences can be  
CC administered to a mammal to raise an immune response, in a method for  
CC interfering with the enzymatic activity of streptococcal  
CC glucosyltransferase in a mammal. The immune response results in reduction  
CC of the colonisation or accumulation of mutans streptococcal strains in  
CC the mammal. Compositions containing W34156 specifically interfere with  
CC the glucan-binding activity of the streptococcal glucosyltransferase.  
CC The peptides can also be used in vaccines for preventing dental caries  
CC in mammals. 22 AA;  
SQ Sequence

Query Match 73.5%; Score 114; DB 25; Length 22;

Best Local Similarity 77.3%; Pred. No. 4.41e-05;  
Matches 17; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Db 1 tgaqtikgklyfkangqykg 22  
| | | | | | | | | | | | | | | | | | | | | |  
QY 1 TGARTINGQLLYFRANGVOYKG 22

RESULT 2  
ID R43696 standard; peptide; 22 AA.  
AC R43696;  
DT 20-MAY-1994 (first entry)  
DE GT subsequence.  
KW GT: glucosyltransferase; vaccine; T-cell; B-cell; reaction;  
KW immunosuppress; peptidyl core matrix; dental caries; diptheria;  
KW tetanus; measles; polio.  
OS Synthetic.  
PN MO932341-A.  
PI 11-NOV-1993.  
PI 30-APR-1993; U04094.  
PR 01-MAY-1992; US-877295.  
PR (FORS-) FORSYTH DENTAL INFIRMARY FOR CHILDREN.  
PI Smith DJ, Taubman MA;  
DR WPI: 93-368721/46.  
PT Synthetic glucosyl-transferase peptide(s) - provoke T- and B-cell  
PT reactions in mammals, and are effective against dental caries  
PS Claim 3; Page 23; 38pp; English.  
CC The sequences (R43694-98) are subsequences from GT, they provoke  
CC T- and B-cell responses in mammals. The can be used to produce a  
CC vaccine comprising 2 of these sequences attached to a peptidyl  
CC core matrix. They are useful in producing T-cell responses to  
CC the GT enzyme causing a reduction of colonisation or accumulation  
CC of mutans streptococcal strains in mammals. The vaccines can be used  
CC in preventing dental caries.  
SQ Sequence 22 AA;

Query Match 73.5%; Score 114; DB 8; Length 22;  
Best Local Similarity 77.3%; Pred. No. 4.41e-05;  
Matches 17; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Db 1 tgaqtikgklyfkangqykg 22  
| | | | | | | | | | | | | | | | | | | | | |  
QY 1 TGARTINGQLLYFRANGVOYKG 22

RESULT 3  
ID W34164 standard; peptide; 24 AA.  
AC W34164;  
DT 18-FEB-1998 (first entry)  
DE GTF antigenic peptide #4 linked to polylysine core.  
KW Glucosyltransferase; GTF; enzyme; immunogen; catalytic domain; vaccine;  
KW surface domain; glucan-binding domain; mutans streptococcal strain;  
KW immune response; glucan-binding activity; dental caries prevention.  
OS Synthetic.  
OS Streptococcus mutans.  
FH Key Location/Qualifiers  
FT Peptide 1..22  
FT /label- GTF antigenic peptide #4 (see W34159)  
FT /note- "attached to the dendritic polylysine core via the  
FT the alpha-amino group of Lys(23); a second copy of  
FT the antigenic 22-mer is linked to Lys(23) via the  
FT omega amino group."  
FT Modified\_site 23  
FT /note- "Lys(23) is linked to one copy of the antigenic  
FT peptide through the alpha-amino group, and to a  
FT second copy of the peptide (not shown) via the  
FT omega amino group."  
FT Modified\_site 24  
FT /note- "the alpha amino acid group of Lys(24) forms a  
FT peptide linkage with the carboxyl amino group of  
FT Lys(23); the omega amino group of Lys(24) forms a  
FT peptide bond with a second Lys residue analogous  
FT to Lys(23)."

PN US568075-A.  
PD 11-NOV-1997.  
PF 01-MAY-1992; 877295.  
PR 30-APR-1993; US-057162.  
PR 01-MAY-1992; US-877295.  
PR (FORS-) FORSYTH DENTAL INFIRMARY FOR CHILDREN.  
PI Smith DJ, Taubman MA;  
DR WPI: 97-558089/51.  
PT Immunogenic compositions containing streptococcal  
PT glucosyl-transferase peptide(s) - used for provoking immune response  
PT to streptococcal glucosyl-transferase for preventing dental caries  
PS Claim 12; Column -; 11pp; English.  
CC W34161-W34165 represent the Streptococcus mutans [glucosyltransferase (GTF)  
CC enzyme immunogenic fragments shown in W34156-W34160 linked to a  
CC polylysine core. W34157 and W34158 are from the catalytic domain of GTF,  
CC while W34160 is from the GTF surface domain. W34156 and W34159 are from  
CC the glucan-binding domain of GTF. These sequences, and the immunogenic  
CC fragments shown in W34156-W34160 can all be used in the immunogenic  
CC composition of the invention. A composition comprising one of these  
CC sequences can be administered to a mammal to raise an immune response, in  
CC a method for interfering with the enzymatic activity of streptococcal  
CC glucosyltransferase in a mammal. The immune response results in reduction  
CC of the colonisation or accumulation of mutans streptococcal strains in  
CC the mammal. Compositions containing W34156 specifically interfere with  
CC the glucan-binding activity of the streptococcal glucosyltransferase.  
CC The peptides can also be used in vaccines for preventing dental caries  
CC in mammals.  
SQ Sequence 24 AA;

Query Match 73.5%; Score 114; DB 25; Length 24;  
Best Local Similarity 77.3%; Pred. No. 4.41e-05;  
Matches 17; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Db 1 tgaqtikgklyfkangqykg 22  
| | | | | | | | | | | | | | | | | | | | | |  
QY 1 TGARTINGQLLYFRANGVOYKG 22

RESULT 4  
ID R32925 standard; Protein; 1592 AA.  
AC R32925;  
DT 28-JUN-1993 (first entry)  
DE Glucosyltransferase I.  
KW GT-1; Streptococcus; dental; caries.  
OS Streptococcus sobrinus.  
PN J05023188-A.  
PN 02-FEB-1993.  
PF 25-JUL-1993; 186592.  
PR 25-JUL-1991; JP-186592.  
PR (FUKU/) FUKUI I.  
PA (KATO/) KATO K.  
DR WPI: 93-079449/10.  
DR N-PSDB: 037760.  
PT DNA sequence glucosyl-transferase-I - comprises Streptococcus  
PT sobrinus DNA sequence with at least one nucleotide added or  
PT deleted.  
PS Claim 13; Page 15; 29pp; Japanese.  
CC The DNA sequence from Streptococcus sobrinus strain 6715 encodes  
CC glucosyltransferase-I (and mutants). The DNA was obtained by treating  
CC S. sobrinus 6715 with mutanolysin, extracting the chromosomal DNA,  
CC partially digesting with Sau3AI and fractionating on agarose gel.  
CC The 3-5 kbp fragment was ligated into pUC18 and E. coli JM109  
CC transformed with it. A GT-1 expressing clone was isolated and  
CC sequenced. The clone may be used in the development of a drug for  
CC dental caries.  
SQ Sequence 1592 AA;

Query Match 72.3%; Score 112; DB 6; Length 1592;  
Best Local Similarity 77.3%; Pred. No. 7.62e-05;  
Matches 17; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Db 1298 tgaqtikgklyfkangqykg 1319  
| | | | | | | | | | | | | | | | | | | | | |



QY 1 TGARTINGQLYFRANGVOYKG 22

RESULT 5  
ID R91047 standard; Protein: 1577 AA.

AC R91047:

DE 22-MAY-1996 (first entry)

Alpha-D-glucosyltransferase.

KW sucrose; transgenic plant; cloning; Escherichia coli;

KM phase lambda-cl13; vector; plasmid pGS6501; plasmid pGS6502;

gene transfer; crop improvement; storage carbohydrate; pasture;

feedstuff; senescence; dextran; binder; food; pharmaceutical.

OS Streptococcus salivarius strain ATCC 25975.

PN WO9606173-A1.

PD 29-FEB-1996.

PF 24-AUG-1995: AU0527.

PR 24-AUG-1994: AU-007643.

PA (GIFP/) GIFPARD P.M.

PA (JACO/) JACOUES N.A.

PA (SIMP/) SIMPSON C.L.

PI Giffard PM, Jacques NA, Simpson CL;

DR MPI: 96-151376/15.

N-PSDB: T13139.

PT Plants contg. new bacterial DNA encoding glucosyl transferase

PT activity - retain higher levels of stored carbohydrate(s) in a form

PS readily digestible by ruminants

PS Claim 4: Page 16-20; 31pp: English.

CC The sequence represents an alpha-D-glucosyltransferase from

CC Streptococcus salivarius. The enzyme is primer-independent, and

CC produces soluble glucan from sucrose. A gene encoding the enzyme

CC may be cloned and expressed in Escherichia coli using a subclone

CC of phase lambda-cl13, e.g. plasmid pGS6501 or plasmid pGS6502. The

CC DNA may also be expressed in a transgenic plant, to improve the

CC level of stored carbohydrate in a pasture plant which normally

CC contains low levels, or to prevent degradation of stored carbohydrate

CC during plant senescence. Dextran may be isolated from the plant, for

CC use as a food binder or pharmaceutical additive. Primer independence

CC ensures that the enzyme will be functional in plants. The glucan is

CC poorly degraded in plants but easily degraded by bacteria in the rumen

CC of grazing livestock.

CC Sequence 1577 AA;

Query Match 56.8%; Score 88; DB 16; Length 1577;

Best Local Similarity 59.1%; Pred. No. 4.64e-02;

Matches 13; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

DB 1533 tggvtingkqlyfdsggrqyk 1554

1 TGARTINGQLYFRANGVOYKG 22

RESULT 6

ID R95014 standard; Protein: 811 AA.

AC R95014:

DE 08-JUL-1996 (first entry)

C. difficile toxin A (aa1870-2680).

KW Toxin A; cytotoxin; enterotoxin; fusion protein; antitoxin;

KM diarrhoea; diagnosis; therapy.

OS Clostridium difficile VPI strain 10463 (ATCC 10463).

PN WO9612802-A1.

PD 02-MAY-1996.

PF 23-OCT-1995: U13737.

PR 24-OCT-1994: US-329154.

PR 16-MAR-1995: US-405496.

PR 14-APR-1995: US-422711.

PR 07-JUN-1995: US-480604.

PA (OPHI-) OPHIDIAN PHARM INC.

PI Filica JR, Kink JA, Padhye NV, Stafford DC, Thalley BS;

PI Williams JA;

DR MPI: 96-230603/23.

PT Fusion proteins comprising non-toxin protein and part of toxin

PT useful to form anti-toxins against Clostridium botulinum type A, and

PT C. difficile type toxins, and to treat C. difficile intoxication,

PT partic. diarrhoea

PS Claim 53: Page 310-312; 434pp: English.

CC A recombinant protein (R95014) comprises amino acids 1870-2680 of

CC Clostridium difficile VPI strain 10463 toxin A (see also R95016),

CC a cytotoxin associated with diarrhoeic disease. It was obt. by

CC PCR amplification of genomic DNA, and was expressed as a soluble

CC fusion protein, with maltose binding protein as fusion partner, in

CC transformed E. coli host cells. The soluble fusion protein can be

CC used to raise avian antibodies useful as antitoxins or diagnostics.

SO Sequence 811 AA;

Query Match 43.2%; Score 67; DB 17; Length 811;

Best Local Similarity 52.2%; Pred. No. 9.21e-00;

Matches 12; Conservative 3; Mismatches 7; Indels 1; Gaps 1;

DB 587 tglrtldgkkyfntntavtvg 609

1 TGARTINGQLYFRAN-GVOYKG 22

RESULT 7

ID R95017 standard; Protein: 812 AA.

AC R95017:

DE 08-JUL-1996 (first entry)

C. difficile toxin A (aa1873-2684).

KW Toxin A; cytotoxin; enterotoxin; fusion protein; antitoxin; vaccine;

KM immunogen; diarrhoea; diagnosis; therapy.

OS Clostridium difficile VPI strain 10463 (ATCC 10463).

PN WO9612802-A1.

PD 02-MAY-1996.

PF 23-OCT-1995: U13737.

PR 24-OCT-1994: US-329154.

PR 16-MAR-1995: US-405496.

PR 14-APR-1995: US-422711.

PR 07-JUN-1995: US-480604.

PA (OPHI-) OPHIDIAN PHARM INC.

PI Filica JR, Kink JA, Padhye NV, Stafford DC, Thalley BS;

PI Williams JA;

DR MPI: 96-230603/23.

PT Fusion proteins comprising non-toxin protein and part of toxin

PT useful to form anti-toxins against Clostridium botulinum type A, and

PT C. difficile type toxins, and to treat C. difficile intoxication,

PT partic. diarrhoea

PS Claim 79: Page 354-356; 434pp: English.

CC A protein (R95017) comprising amino acids 1873-2684 of Clostridium

CC difficile VPI strain 10463 toxin A (see also R95016) was obt. by

CC PCR amplification of genomic DNA. The PCR product can be expressed

CC in transformed E. coli host cells as a soluble fusion protein, with

CC maltose binding protein or a polystyrene affinity tag as

CC fusion partner. The resulting soluble fusion proteins, which are

CC substantially endotoxin-free, can be used in anti-clostridial

CC vaccines or to raise avian antibodies useful in clostridial antitoxin

CC therapy for humans and animals.

SO Sequence 812 AA;

Query Match 43.2%; Score 67; DB 17; Length 812;

Best Local Similarity 52.2%; Pred. No. 9.21e-00;

Matches 12; Conservative 3; Mismatches 7; Indels 1; Gaps 1;

DB 588 tglrtldgkkyfntntavtvg 610

1 TGARTINGQLYFRAN-GVOYKG 22

RESULT 8

ID R95016 standard; Protein: 2710 AA.

AC R95016:

DE 08-JUL-1996 (first entry)

C. difficile toxin A.

KW Toxin A; cytotoxin; enterotoxin; fusion protein; antitoxin;

KM vaccine; diarrhoea; diagnosis; therapy.

OS Clostridium difficile VPI strain 10463 (ATCC 10463).

Query Match	43.28;	Score 67;	DB 17;	Length 2710;
Best Local Similarity	52.28;	Pred. No. 9.21e+00;		
Matches	12;	Conservative	3;	Mismatches 7; Indels 1; Gaps 1
Db	2460	tgllrlgdkkkyfntntavrtg	2482	
Oy	1	tgartingcollfrran-gvovkg	22	

RESULT 9  
 ID M68387 standard; Protein; 2710 AA.  
 AC W68387;  
 DT 07-DEC-1998 (first entry)  
 DE Clostridium difficile toxin A.  
 KW Antitoxin; vaccine; cytotoxin; toxin A; intoxication; immunogen;  
 KN pseudomembranous enterocolitis.  
 OS Clostridium difficile.  
 PN M09808540-A1.  
 PR 05-MAR-1998.  
 PS 28-AUG-1997; U15394.  
 PT 28-AUG-1997; US-704159.  
 PA (OPHI-) OPHIDIAN PHARM INC.  
 PI Thalley BS, Williams JN;  
 DR WPI: 98-230234/20.  
 DR N-PSDB: V30560.  
 PT Host cell containing recombinant expression vector encoding  
 PT Clostridium botulinum type B or E toxin - useful to treat humans  
 PT and other animals at risk of intoxication with clostridial toxin  
 PS Example 15: Page 220-230: 428pp; English.  
 CC This is the amino acid sequence of Clostridium difficile toxin A,  
 CC deduced from the coding region (see V30560) of the toxin A gene.  
 CC Toxin A is a potent cytotoxin that plays a direct role in damaging  
 CC gastrointestinal tissues. Severe cases of C. difficile  
 CC intoxication result in pseudomembranous colitis. This would be  
 CC prevented by neutralising the effects of toxin A in the  
 CC gastrointestinal tract. Examples are provided of the production  
 CC of recombinant C. difficile toxin A in host cells and of the in  
 CC vivo neutralisation of toxin A by antibodies against recombinant  
 CC toxin A polypeptides. The invention specifically relates to  
 CC recombinant proteins derived from Clostridium botulinum toxins  
 CC (see M68389-400) and their use as immunogens for the production of  
 CC vaccines and antitoxins.  
 SO Sequence 2710 AA.

Query Match 43.28; Score 67; DB 35; Length 2710;

RESULT	10
ID	R78621
	standard; Protein; 404 AA

RESULT	10	
ID	R78621	standard; Protein; 404 AA.
AC	R78621;	
DT	11-JUN-1996	(first entry)
DE	Chicken GalNAc-alpha-2,6-sialyltransferase P-B3.	
KW	GalNAc-alpha-2,6-sialyltransferase P-B3; anti-inflammatory;	
KM	viral infection; tumour; migration; inhibitor; sugar chain;	
KW	reagent; chicken.	
OS	Gallus gallus.	
PN	WO9518217-A1.	
PD	06-JUL-1995.	
PF	22-DEC-1994; J02182.	
PR	24-DEC-1993; JP-348250.	
PR	28-MAR-1994; JP-057369.	
PR	28-APR-1994; JP-091507.	
PA	(RIKA ) INST PHYSICAL & CHEM RES.	
PI	Hamamoto T, Kojima N, Kurosawa N, Lee Y, Nakooka T;	
PI	Tsuyji S;	
DR	WPI; 95-246383/32.	
DR	N-PSDB; Q95325.	
PT	New GalNAc-alpha-2,6-sialyl:transferases P-B1 and P-B3 - for e.g	
PT	treatment of genetic disorders involving missing sugars	
CS	Claim 4; Pages 52-55; 70pp; Japanese.	
CC	Q95325 encodes R78621 chicken GalNAc-alpha-2,6-sialyltransferase	
CC	P-B3. P-B3 can be used as a reagent for introducing human type	
CC	sugar chains onto proteins, or for the treatment of genetic	
CC	disorders involving missing sugar chains. It may also be used as	
CC	a tumour migration inhibitor, viral infection preventative and	
CC	as an anti-inflammatory.	
QO	Sequence 404 AA;	

Query Match	40.0%	Score 63	DB 16	Length 404
Best Local Similarity	45.0%	Pred. No.	2.40e+05	
Matches	9	Conservative	5	Mismatches 6
				Indels 0
				Gaps 0
Db	196	graldaholvirringaltkg	215	
		..: : ::		
Oy	3	ARTINGOLLYRBRANGVOYKG	22	

RESULT 11  
ID M06416 standard; protein; 349 AA.  
AC M06416;  
DT 28-JAN-1997 (first entry)  
DE phospholipase-related protein.  
KW phospholipase-related protein; mouse; PRP; recessive cpk; murine;  
KW polycystic kidney disease; PKD; acute renal injury; autosomal dominant;  
KW autosomal recessive; fluid-filled cyst; nephron; collecting duct; kidney  
KW renal parenchyma; renal failure; mammal; diagnostic marker; therapy;  
KW acute oryphosphate toxicity; phospholipase inhibitor;  
KW anti-cancer drug enhancer; anti-HIV drug enhancer.  
OS Mus musculus.  
PN US5552313-A.  
PD 03-SEP-1996.  
PF 21-NOV-1994; 343027.  
PR 21-NOV-1994; US-343027.  
PA (UNIV ) UNIV KANSAS.  
PI Calvet JP, Hou X, Magenheimer BS, Maser RL;  
DR WPI; 96-412067/41.  
DR N-PSDS; T43210.  
PT Mouse phospho:tri:esterase-related protein DNA - used to develop  
PT prods. for diagnosis of poly:cystic kidney disease or acute renal  
PT failure  
PS Claim 1; Column 25-28; 23pp; English.  
CC This sequence represents the mouse phospholipase related protein

(mPRP). The mPRP sequence has abnormal underexpression in the recessive cpx murine model of a polycystic kidney disease (PKD), and has decreased expression following acute renal injury. PKD is a common disease (affecting 1 in 500-1000 individuals) and can be inherited as an autosomal dominant or as an autosomal recessive. PKD is characterised by the development of innumerable, large, fluid-filled cysts from the nephrons and collecting ducts of affected kidneys. It is thought that enlargement of the cysts interferes with functioning of the normal renal parenchyma, which eventually leads to renal failure. The mPRP, or other similar mammalian PRPs can be used as diagnostic markers for PKD and acute renal failure. The PRPs can also be used in therapeutic and protective treatments, such as for acute organophosphate toxicity. They can also be used to develop products to enhance the effectiveness of other types of therapy, e.g. phosphodiesterase inhibitors to enhance the effectiveness of certain anti-cancer, or anti-HIV drugs.

Query Match 40.08; Score 62; DB 19; Length 349;  
Best Local Similarity 53.88; Pred. No. 3.04e+01;  
Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

79 atreellyfkayg 91  
: | |||||: |  
QY 5 TINGOLLYFRANG 17

RESULT 12  
ID W38549 standard; Protein: 207 AA.  
AC W38549;  
DE 06-NOV-1998 (first entry)  
KW Streptococcus pneumoniae protein; genetic immunisation; antagonist;  
KW immunological response; inoculation; antibody production; inhibitor;  
KW T cell immune response; antimicrobial compound; bacterial adhesion;  
KW extracellular matrix protein; protein-mediated cell invasion; wound;  
KW pathogenesis.  
OS Streptococcus pneumoniae.  
PN W09743303-A1.  
PD 20-NOV-1997.  
PF 14-MAY-1997; U07950.  
PI 14-MAY-1996; US-017670.  
PA (SMK ) SMITHKLINE BEECHAM CORP.  
PI (SMK ) SMITHKLINE BEECHAM PLC.  
PI Black MT, Hodgson JE, Knowles DJC, Nicholas RO,  
PI Stodola RK;  
PI WPI: 98-008793/01.  
DR N-PSDB: T98605.  
PT Novel Streptococcus pneumoniae proteins and related DNA - useful for  
PT diagnosing anti-microbial agents for treatment of bacterial  
PT infections.  
BT Claim 12; Page 327; 483pp; English.  
AB This sequence represents a Streptococcus pneumoniae protein that, based  
CC on homology with Bacteroides nodosus protein, is a peptide chain release  
CC factor 3 (RF-3), and is encoded by a DNA sequence of the invention.  
CC The DNA sequences were isolated from Streptococcus pneumoniae strain  
CC 0100993 (NCIMB 40794). The Streptococcus pneumoniae proteins of the  
CC invention can be used to identify compounds which interact with and  
CC inhibit or activate the activity of the proteins. Antagonists can be  
CC used to treat diseases caused by S. pneumoniae proteins, through genetic  
CC immunisation. They can also be used to induce an immunological response  
CC in a mammal by inoculation with the S. pneumoniae proteins or delivery  
CC and/or T cell immune responses to protect the animal from disease. The  
CC proteins can also be used to identify antimicrobial compounds which are  
CC capable of inhibiting their bioactivity. In particular the proteins of  
CC the invention can be used to prevent adhesion of bacteria to mammalian  
CC extracellular matrix proteins on in-dwelling devices or in wounds, to  
CC block protein-mediated mammalian cell invasion, and to block the normal  
CC progression of pathogenesis in infections initiated other than by the  
CC implantation of in-dwelling devices or other surgical techniques.  
SQ Sequence 207 AA;

Query Match 39.48; Score 61; DB 34; Length 207;

Best Local Similarity 61.58; Pred. No. 3.85e+01;  
Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;  
DB 21 agttiteqlylf 33  
: | || |||||  
QY 1 TGARTINGOLLYF 13

RESULT 13  
ID W83359 standard; Protein: 514 AA.  
AC W83359;  
DE 17-FEB-1999 (first entry)  
KW Streptococcus pneumoniae prfC protein sequence.  
KW Streptococcus pneumoniae; prfC; antibacterial; meningitis; pneumonia;  
KW peptide releasing factor; otitis media; conjunctivitis; bacteraemia;  
KW sinusitis; pleural empyema; endocarditis; immunisation; infection;  
KW gene therapy.  
OS Streptococcus pneumoniae.  
PN EP-881292-A2.  
PD 02-DEC-1998.  
PF 26-MAY-1998; 304157.  
PI 29-MAY-1997; US-865311.  
PA (SMK ) SMITHKLINE BEECHAM CORP.  
PI Pearson SC;  
PI WPI: 99-001398/01.  
DR N-PSDB: V72588.  
PT New Streptococcus pneumoniae peptide releasing factor polypeptide  
PT and polynucleotide - useful as diagnostic reagents and for  
PT prevention and treatment of diseases caused by bacterial infections,  
PT including meningitis and pneumonia.  
BT Claim 12; Page 24-26; 27pp; English.  
AB The present sequence is Streptococcus pneumoniae prfC gene useful for  
CC diagnosing susceptibility to diseases and polynucleotides are useful for  
CC polymorphisms of the prfC gene. PCR using prfC probes is useful for  
CC diagnosing diseases caused by organisms comprising the prfC gene by  
CC detection at the nucleic acid level, and analysing for the presence or  
CC amount of prfC polypeptide in cell or tissue samples. This method is  
CC useful for diagnosing the stage of infection and the type of pathogen.  
CC prfC polypeptides and polynucleotides can be used to screen for  
CC antagonists and agonists (especially bacteriostatic and bacteriocidal  
CC compounds), which can be used in treatment to enhance or block prfC  
CC activity. prfC polypeptide is useful for screening for antibacterial  
CC compounds which can be used as drugs. prfC polynucleotides can be used  
CC in genetic immunisation (gene therapy) to protect against bacterial  
CC infections. prfC polypeptides, polynucleotides and antagonists may be  
CC used as a wound treatment to prevent adhesion of bacteria to matrix  
CC proteins, as they interfere with the physical interaction between the  
CC pathogen and mammalian host. prfC antibodies are also useful for inducing  
CC an immune response to immunise and prevent disease, and for isolating  
CC prfC clones or purifying the peptide by affinity chromatography. Diseases  
CC diagnosed, prevented or treated include: otitis media, conjunctivitis,  
CC pneumonia, bacteraemia, sinusitis, pleural empyema, endocarditis and  
CC especially meningitis.  
SQ Sequence 514 AA;

Query Match 39.48; Score 61; DB 39; Length 514;  
Best Local Similarity 61.58; Pred. No. 3.85e+01;  
Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;  
DB 21 agttiteqlylf 33  
: | || |||||  
QY 1 TGARTINGOLLYF 13

RESULT 14  
ID W03642 standard; peptide: 116 AA.  
AC W03642;  
DE 17-DEC-1996 (first entry)  
KW Human cannabinoid GPR N-terminal sequence.  
KW G-protein coupled receptor; ligand binding assay; transmembrane domain;  
KW schizophrenia; dopamine; cAMP; adenosine; thrombin; adrenergic; opsin;  
KW muscarinic acetylcholine; endothelin; bombesin; endocrine; rhodopsin;

Search completed: Tue Jan 11 15:46:47 2000  
 Job time : 19 secs.

KW Odorant; cytomegalovirus; serotonergic.  
 OS Homo sapiens.  
 PN US5508384-A.  
 PD 16-APR-1996.  
 PF 10-SEP-1992: 943236.  
 PR 10-SEP-1992: US-943236.  
 PR 09-SEP-1993: US-118270.  
 PA (UYN ) UNIV NEW YORK STATE.  
 PI Murphy RB, Schuster DI;  
 DR WPI: 96-208785/21.  
 PT New dopamine receptor peptide - useful as antipsychotic agent, e.g.  
 PT for treating schizophrenia  
 PS Disclosure; Fig 8B(3); 184pp: English.  
 CC Proteins W02657-W02720 represent a range of G-protein coupled receptor  
 (GPR) proteins selected from cAMP, adenosine, muscarinic acetylcholine,  
 adrenergic, thrombin, endothelin, bombesin, endocrine, rhodopsin, opsin,  
 odorant, cytomegaloviral and other GPR proteins. The peptides  
 W03578-W03651 represent the N-terminal fragments of the above proteins.  
 CC The receptor proteins were used to design polypeptides, pref. based on  
 the transmembrane domains, for use in G-protein coupled receptor ligand  
 binding assays. The polypeptide fragments retain biological activity  
 CC such as binding a GPR ligand or modulating GPR ligand binding to a GPR  
 (see W02747-W02999 for examples of polypeptide fragments).  
 CC The polypeptide fragments can be used in compositions for treating  
 CC subjects suffering from a pathology related to a GPR abnormality e.g. a  
 CC psychotic disorder such as schizophrenia.  
 SQ Sequence 116 AA;

## Query Match

Best Local Similarity 38.7%; Score 60; DB 19; Length 116;  
 Best Local Similarity 50.0%; Pred. No. 4.87e+01;

Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Db 14 rttldillygsndiq 29  
 |||:||||:|:|  
 QY 4 RTINGQLLYFRANGVQ 19  
 |

## RESULT 15

ID R14196 standard; Protein; 472 AA.  
 AC R14196;  
 DT 19-DEC-1991 (first entry)  
 DE Human cannabinoid receptor.  
 KW Cannabis sativa; marijuana; drug test; substance K receptor.  
 OS Homo sapiens.  
 PN US7564075-A.  
 PD 03-SEP-1991.  
 PF 08-AUG-1990: 564075.  
 PF 08-AUG-1990: US-564075.  
 F (USSH ) NAT INST OF HEALTH.  
 PI Matsuda L, Brownstein M, Bonner T;  
 DR WPI: 91-303326/41.  
 DR N-PSDB: Q14003.  
 PT DNA encoding mammalian cannabinoid receptor - used for producing  
 PT receptor for screening drugs and ligands and in detection  
 PS Disclosure; Fig 5; 25pp: English.  
 CC SKR6 cDNA encoding the rat cannabinoid receptor was isolated from a  
 CC rat cerebral cortex cDNA library. An EcoRI-XbaI fragment was used to  
 CC screen a human cosmid library. A positive clone was identified and  
 CC sequenced. There is ca. 97 per cent homology between this amino  
 CC acid sequence deduced from the sequence of the positive human clone  
 CC and rat cannabinoid receptor (see R14195). Recombinantly produced  
 CC receptor can be used to screen for new drugs suitable for treatment  
 CC of cannabinoid-treatable conditions, e.g. glaucoma, bronchial asthma,  
 CC etc.  
 SQ Sequence 472 AA;

## Query Match

Best Local Similarity 38.7%; Score 60; DB 3; Length 472;  
 Best Local Similarity 50.0%; Pred. No. 4.87e+01;

Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Db 14 rttldillygsndiq 29  
 |||:||||:|:|  
 QY 4 RTINGQLLYFRANGVQ 19

\*\*\*\*\*  
 WIRELESS (TM)  
 \*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
 Copyright (c) 1993-1998 University of Edinburgh, U.K.  
 Distribution rights by Oxford Molecular Ltd

MPsrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Tue Jan 11 15:49:24 2000; MasPar time 1.77 Seconds

lar output not generated. 147.986 Million cell updates/sec

Title: >US-09-290-049-3  
 Description: (1-22) from US09290049.pep  
 Perfect score: 155  
 Sequence: 1 TGARTINGQLLYFRANGVQVKG 22

Scoring table: PAM 150  
 Gap 15

Searched: 122461 seqs, 11912985 residues

Post-processing: Minimum Match 0%  
 Listing first 45 summaries

Database: a-issued  
 1:5A\_COMB 2:5B\_COMB 3:PCT9\_COMB 4:backfiles1

Statistics: Mean 19.644; Variance 62.865; scale 0.312

Prod. No. is the number of results predicted by chance to have a  
 score greater than or equal to the score of the result being printed,  
 and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Prod. No.
1	114	73.5	22	1	US-08-057-Sequence 3, Applicatio	2.93e+05
2	67	43.2	811	2	US-08-405-Sequence 7, Applicatio	5.11e+00
3	67	43.2	811	1	US-08-480-Sequence 7, Applicatio	5.11e+00
4	67	43.2	812	1	US-08-480-Sequence 29, Applicati	5.11e+00
5	67	43.2	2710	2	US-08-405-Sequence 6, Applicatio	5.11e+00
6	67	43.2	2710	1	US-08-480-Sequence 6, Applicatio	5.11e+00
7	63	40.6	404	2	US-08-666-Sequence 7, Applicatio	1.32e+01
8	62	40.0	349	1	US-08-343-Sequence 12, Applicati	1.67e+01
9	61	39.4	514	2	US-08-865-Sequence 2, Applicatio	2.11e+01
10	59	38.1	91	1	US-08-480-Sequence 8, Applicatio	3.36e+01
11	59	38.1	91	2	US-08-405-Sequence 8, Applicatio	3.36e+01
12	59	38.1	453	2	US-08-485-Sequence 8, Applicatio	3.36e+01
13	59	38.1	453	2	US-08-422-Sequence 8, Applicatio	3.36e+01
14	59	38.1	453	2	US-08-476-Sequence 6, Applicatio	3.36e+01
15	59	38.1	453	2	US-08-956-Sequence 6, Applicatio	3.36e+01
16	59	38.1	453	2	US-08-611-Sequence 8, Applicatio	3.36e+01
17	59	38.1	998	3	PCT-US96-0-Sequence 2, Applicatio	3.36e+01
18	59	38.1	998	3	PCT-US96-0-Sequence 2, Applicatio	3.36e+01
19	59	38.1	998	3	PCT-US96-0-Sequence 17, Applicati	4.32e+01
20	58	37.4	260	2	US-07-857-Sequence 27, Applicati	5.32e+01
21	57	36.8	363	4	US-08-001-Patent No. 522306	5.32e+01
22	57	36.8	488	1	US-08-001-Sequence 2, Applicatio	5.32e+01
23	57	36.8	488	1	US-07-794-Sequence 2, Applicatio	5.32e+01

RESULT ID	Sequence	Score	Prod. No.	US-08-057-162B-3	STANDARD	PRT	22 AA.
24	57	36.8	492	1	US-07-794-Sequence 4, Applicatio	5.32e+01	
25	57	36.8	492	1	US-08-001-Sequence 4, Applicatio	5.32e+01	
26	57	36.8	501	4	5168064-4-Patent No. 5168064	5.32e+01	
27	57	36.8	501	1	US-08-271-Sequence 2, Applicatio	5.32e+01	
28	57	36.8	501	1	US-07-687-Sequence 2, Applicatio	5.32e+01	
29	57	36.8	501	1	US-08-434-Sequence 2, Applicatio	5.32e+01	
30	56	36.1	15	4	5439796-5-Patent No. 5439796	6.68e+01	
31	56	36.1	54	2	US-08-797-Sequence 23, Applicatio	6.68e+01	
32	56	36.1	448	2	US-08-231-Sequence 2, Applicatio	6.68e+01	
33	56	36.1	506	2	US-08-635-Sequence 2, Applicatio	6.68e+01	
34	55	35.5	265	2	US-09-002-Sequence 2, Applicatio	8.38e+01	
35	55	35.5	265	1	US-08-461-Sequence 2, Applicatio	8.38e+01	
36	55	35.5	265	1	PCT-US94-1-Sequence 3, Applicatio	8.38e+01	
37	55	35.5	274	1	US-08-287-Sequence 3, Applicatio	8.38e+01	
38	55	35.5	448	2	US-08-074-Sequence 3, Applicatio	8.38e+01	
39	55	35.5	448	3	PCT-US94-0-Sequence 3, Applicatio	8.38e+01	
40	55	35.5	971	1	US-08-446-Sequence 19, Applicati	8.38e+01	
41	55	35.5	971	1	US-09-066-Sequence 19, Applicati	8.38e+01	
42	55	35.5	971	1	US-08-446-Sequence 19, Applicati	8.38e+01	
43	55	35.5	971	2	US-08-805-Sequence 19, Applicati	8.38e+01	
44	55	35.5	971	2	US-08-064-Sequence 19, Applicati	8.38e+01	
45	55	35.5	993	1	US-08-446-Sequence 25, Applicati	8.38e+01	

Sequence 3, Application US/08057162B

Sequence 3, Application US/08057162B

Patent No. 5686075

GENERAL INFORMATION:

APPLICANT: Taubman, Martin A.

INVENTOR: Smith, Daniel J.

TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINE FOR DENTAL CARRIES

NUMBER OF SEQUENCES: 5

CORRESPONDENCE ADDRESS:

ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.

STREET: Two Millitia Drive

CITY: Lexington

STATE: Massachusetts

COUNTRY: USA

ZIP: 02173

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/057.162B

FILING DATE: 30-APR-1993

CLASSIFICATION: 424

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/877.295

FILING DATE: 01-MAY-1992

ATTORNEY/AGENT INFORMATION:

NAME: Wagner, Richard W.

REGISTRATION NUMBER: 34,480

REFERENCE/DOCKET NUMBER: FDC92-01A

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617) 861-6240

TELEFAX: (617) 861-9540

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 22 amino acids

TYPE: amino acid

STRANDEDNESS: single

CC TOPOLOGY: 1linear  
CC MOLECULE TYPE: peptide  
SQ SEQUENCE 22 AA; 2366 MW; 2475 CN;  
Query Match 73.5%; Score 114; DB 1; Length 22;  
Best Local Similarity 77.3%; Pred. No. 2.99e-05;  
Matches 17; Conservative 3; Mismatches 2; Indels 0; Gaps 0;  
DB 1 TGAQTIGKGLYFRANGOVKG 22  
1 TGARTINGQLLYFRANGOVKG 22  
OY  
RESULT 2  
ID US-08-405-496A-7 STANDARD; PRT; 811 AA.  
XX  
AC xxxxxx  
XX  
DE  
Sequence 7, Application US/08405496A  
CC Sequence 7, Application US/08405496A  
CC Patent No. 5919665  
CC GENERAL INFORMATION:  
CC APPLICANT: WILLIAMS, JAMES A.  
CC TITLE OF INVENTION: VACCINE FOR CLOSTRIDIUM BOTULINUM  
CC TITLE OF INVENTION: NEUROTOXIN  
CC NUMBER OF SEQUENCES: 30  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: MEDLEN & CARROLL, LLP  
CC STREET: 220 MONTGOMERY STREET, SUITE 2200  
CC CITY: SAN FRANCISCO  
CC STATE: CALIFORNIA  
CC COUNTRY: USA  
CC ZIP: 94104  
CC  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: Patent In Release #1.0, Version #1.30  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/405,496A  
CC FILING DATE: 16-MAR-1995  
CC CLASSIFICATION: 424  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 08/329,154  
CC FILING DATE: 25-OCT-1994  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 08/161,907  
CC FILING DATE: 02-DEC-1993  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 07/985,321  
CC FILING DATE: 04-DEC-1992  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 07/429,791  
CC FILING DATE: 31-OCT-1989  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: INGOLIA, DIANE E.  
CC REGISTRATION NUMBER: 40,027  
CC REFERENCE/DOCKET NUMBER: OPHD-01308  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: (415) 705-8410  
CC TELEFAX: (415) 397-8338  
CC INFORMATION FOR SEQ ID NO: 7:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 811 amino acids  
CC TYPE: amino acid  
CC STRANDEDNESS: unknown  
CC TOPOLOGY: unknown  
CC MOLECULE TYPE: protein  
CC SEQUENCE 811 AA; 91921 MW; 3356332 CN;

Query Match 43.2%; Score 67; DB 2; Length 811;  
Best Local Similarity 52.2%; Pred. No. 5.11e-00;  
Matches 12; Conservative 3; Mismatches 7; Indels 1; Gaps 1;  
DB 587 TGLRTIDGKKYYNTNTAVAVTG 609  
1 TGARTINGQLLYFRAN-GVOVKG 22  
OY  
RESULT 3  
ID US-08-480-604A-7 STANDARD; PRT; 811 AA.  
XX  
AC xxxxxx  
XX  
DE  
Sequence 7, Application US/08480604A  
CC Sequence 7, Application US/08480604A  
CC Patent No. 5736139  
CC GENERAL INFORMATION:  
CC APPLICANT: KINK, JOHN A.  
CC APPLICANT: THALLEY, BRUCE S.  
CC APPLICANT: PADRYE, NISRA V.  
CC APPLICANT: FURCA, JOSEPH R.  
CC APPLICANT: STAFFORD, DOUGLAS C.  
CC TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND  
CC TITLE OF INVENTION: PREVENTION OF C. DIFFICILE DISEASE  
CC NUMBER OF SEQUENCES: 32  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: MEDLEN & CARROLL, LLP  
CC STREET: 220 MONTGOMERY STREET, SUITE 2200  
CC CITY: SAN FRANCISCO  
CC STATE: CALIFORNIA  
CC COUNTRY: UNITED STATES OF AMERICA  
CC ZIP: 94104  
CC  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: Patent In Release #1.0, Version #1.30  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/480,604A  
CC FILING DATE: 07-JUN-1995  
CC CLASSIFICATION: 424  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 08/422,711  
CC FILING DATE: 14-APR-1995  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 08/405,496  
CC FILING DATE: 16-MAR-1995  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 08/329,154  
CC FILING DATE: 25-OCT-1994  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 08/161,907  
CC FILING DATE: 02-DEC-1993  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 07/985,321  
CC FILING DATE: 04-DEC-1992  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 07/429,791  
CC FILING DATE: 31-OCT-1989  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: INGOLIA, DIANE E.  
CC REGISTRATION NUMBER: 40,027  
CC REFERENCE/DOCKET NUMBER: OPHD-01763  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: (415) 705-8410  
CC TELEFAX: (415) 397-8338  
CC INFORMATION FOR SEQ ID NO: 7:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 811 amino acids

CC TYPE: amino acid  
CC STRANDEDNESS: unknown  
CC TOPOLOGY: unknown  
CC MOLECULE TYPE: Protein  
SQ SEQUENCE 811 AA; 91921 MW; 3356332 CN;  
Query Match 43.2%; Score 67; DB 1; Length 811;  
Best Local Similarity 52.2%; Pred. No. 5,11e+00;  
Matches 12; Conservative 3; Mismatches 7; Indels 1; Gaps 1;  
Db 587 TGLRTIDGKKYENTNAYAVTG 609  
QY 1 TGARTINGOLLYFRAN-GVOYKG 22  
RESULT 4  
ID US-08-480-604A-29 STANDARD; PRT; 812 AA.  
XX  
AC xxxxxx  
XX  
Sequence 29, Application US/08480604A  
XX Patent No. 5736139  
CC GENERAL INFORMATION:  
CC APPLICANT: KINK, JOHN A.  
CC APPLICANT: THALLEY, BRUCE S.  
CC APPLICANT: PADHAY, NISHA V.  
CC APPLICANT: FIRCA, JOSEPH R.  
CC APPLICANT: STAFFORD, DOUGLAS C.  
CC TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND  
CC PREVENTION OF C. DIFFICILE DISEASE  
CC NUMBER OF SEQUENCES: 32  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: MEDLEN & CARROLL, LLP  
CC STREET: 220 MONTGOMERY STREET, SUITE 2200  
CC CITY: SAN FRANCISCO  
CC STATE: CALIFORNIA  
CC COUNTRY: UNITED STATES OF AMERICA  
CC ZIP: 94104  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: Patent Release #1.0, Version #1.30  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/480,604A  
CC FILING DATE: 07-JUN-1995  
CC CLASSIFICATION: 424  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 08/422,711  
CC FILING DATE: 14-APR-1995  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 08/405,496  
CC FILING DATE: 16-MAR-1995  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 08/329,154  
CC FILING DATE: 25-OCT-1994  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 08/161,907  
CC FILING DATE: 02-DEC-1993  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 07/985,321  
CC FILING DATE: 04-DEC-1992  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 07/429,791  
CC FILING DATE: 31-OCT-1989  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: INGOLIA, DIANE E.  
CC REGISTRATION NUMBER: 40,027  
CC REFERENCE/DOCKET NUMBER: OPD-01763

CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: (415) 705-8410  
CC TELEFAX: (415) 397-8338  
CC INFORMATION FOR SEQ ID NO: 29:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 812 amino acids  
CC TYPE: amino acid  
CC STRANDEDNESS: unknown  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: Protein  
SQ SEQUENCE 812 AA; 92022 MW; 3364670 CN;  
Query Match 43.2%; Score 67; DB 1; Length 812;  
Best Local Similarity 52.2%; Pred. No. 5,11e+00;  
Matches 12; Conservative 3; Mismatches 7; Indels 1; Gaps 1;  
Db 588 TGLRTIDGKKYENTNAYAVTG 610  
QY 1 TGARTINGOLLYFRAN-GVOYKG 22  
RESULT 5  
ID US-08-405-496A-6 STANDARD; PRT; 2710 AA.  
XX  
AC xxxxxx  
XX  
Sequence 6, Application US/08405496A  
XX Patent No. 5919665  
CC GENERAL INFORMATION:  
CC APPLICANT: WILLIAMS, JAMES A.  
CC TITLE OF INVENTION: VACCINE FOR CLOSTRIDIUM BOTULINUM  
CC NUMBER OF SEQUENCES: 30  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: MEDLEN & CARROLL, LLP  
CC STREET: 220 MONTGOMERY STREET, SUITE 2200  
CC CITY: SAN FRANCISCO  
CC STATE: CALIFORNIA  
CC COUNTRY: USA  
CC ZIP: 94104  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: Patent Release #1.0, Version #1.30  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/405,496A  
CC FILING DATE: 16-MAR-1995  
CC CLASSIFICATION: 424  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 08/329,154  
CC FILING DATE: 25-OCT-1994  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 08/161,907  
CC FILING DATE: 02-DEC-1993  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 07/985,321  
CC FILING DATE: 04-DEC-1992  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 07/429,791  
CC FILING DATE: 31-OCT-1989  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: INGOLIA, DIANE E.  
CC REGISTRATION NUMBER: 40,027  
CC REFERENCE/DOCKET NUMBER: OPD-01308  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: (415) 705-8410  
CC TELEFAX: (415) 397-8338  
CC INFORMATION FOR SEQ ID NO: 6:

CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 2710 amino acids  
CC TYPE: amino acid  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: protein  
SQ SEQUENCE 2710 AA: 308053 MW: 38468929 CN;

Query Match 43.28; Score 67; DB 2; Length 2710;  
Best Local Similarity 52.28; Pred. No. 5.11e+00;  
Matches 12; Conservative 3; Mismatches 7; Indels 1; Gaps 1;

Db 2460 TGLRTIDGKKYFNTAVATG 2482  
QY 1 TGARTINGQLIFRAN-GVOYKG 22

RESULT 6  
ID US-08-480-604A-6 STANDARD; PRT: 2710 AA.  
MA xxxxxx

Sequence 6, Application US/08480604A  
Patent No. 5736139  
GENERAL INFORMATION:  
CC APPLICANT: KIRK, JOHN A.  
CC APPLICANT: THALLEY, BRUCE S.  
CC APPLICANT: PADHYE, NISHA V.  
CC APPLICANT: FIRCA, JOSEPH R.  
CC APPLICANT: STAFFORD, DOUGLAS C.  
CC TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND  
CC TITLE OF INVENTION: PREVENTION OF C. DIFFICILE DISEASE  
CC NUMBER OF SEQUENCES: 32  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: MEDLEN & CARROLL, LLP  
CC STREET: 220 MONTGOMERY STREET, SUITE 2200  
CC CITY: SAN FRANCISCO  
CC STATE: CALIFORNIA  
CC COUNTRY: UNITED STATES OF AMERICA  
CC ZIP: 94104  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: Patentin Release #1.0, Version #1.30  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/480,604A  
CC FILING DATE: 07-JUN-1995  
CC CLASSIFICATION: 424  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 08/422,711  
CC FILING DATE: 14-APR-1995  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 08/405,496  
CC FILING DATE: 16-MAR-1995  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 08/329,154  
CC FILING DATE: 25-OCT-1994  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 08/161,907  
CC FILING DATE: 02-DEC-1993  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 07/985,321  
CC FILING DATE: 04-DEC-1992  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 07/429,791  
CC FILING DATE: 31-OCT-1989  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: INGOLIA, DIANE E.  
CC REGISTRATION NUMBER: 40,027

CC REFERENCE/DOCKET NUMBER: OPFD-01763  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: (415) 705-8410  
CC TELEFAX: (415) 397-8338  
CC INFORMATION FOR SEQ ID NO: 6:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 2710 amino acids  
CC TYPE: amino acid  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: protein  
SQ SEQUENCE 2710 AA: 308053 MW: 38468929 CN;

Query Match 43.28; Score 67; DB 1; Length 2710;  
Best Local Similarity 52.28; Pred. No. 5.11e+00;  
Matches 12; Conservative 3; Mismatches 7; Indels 1; Gaps 1;

Db 2460 TGLRTIDGKKYFNTAVATG 2482  
QY 1 TGARTINGQLIFRAN-GVOYKG 22

RESULT 7  
ID US-08-666-367B-7 STANDARD; PRT: 404 AA.  
MA xxxxxx

Sequence 7, Application US/08666367B  
Patent No. 5854042  
GENERAL INFORMATION:  
CC APPLICANT: Shuichi TSUJI et al.  
CC TITLE OF INVENTION: NOVEL SUGAR-CHAIN SYNTHETASE AND PROCESS FOR  
CC TITLE OF INVENTION: PRODUCING THE SAME  
CC NUMBER OF SEQUENCES: 8  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: Wenderoth, Lind & Ponack  
CC STREET: 805 Fifteenth Street, N.W., #700  
CC CITY: Washington  
CC STATE: D.C.  
CC COUNTRY: U.S.A.  
CC ZIP: 20005  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb  
CC COMPUTER: IBM compatible  
CC OPERATING SYSTEM: MS-DOS  
CC SOFTWARE: Wordperfect 5.1  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/666,367B  
CC FILING DATE: August 19, 1996  
CC CLASSIFICATION: 435  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER:  
CC FILING DATE:  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Warren M. Cheek, Jr.  
CC REGISTRATION NUMBER: 33,367  
CC REFERENCE/DOCKET NUMBER:  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: 202-371-8850  
CC TELEFAX:  
CC TELEX:  
CC INFORMATION FOR SEQ ID NO: 7:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 404 amino acids  
CC TYPE: amino acid  
CC STRANDEDNESS: single  
CC TOPOLOGY: linear  
CC ORIGINAL SOURCE:  
CC ORGANISM: G. gallus (chicken)  
SQ SEQUENCE 404 AA: 45826 MW: 844143 CN;





CC APPLICANT: THALLEY, BRUCE S.  
CC APPLICANT: PADHYE, NISHA V.  
CC APPLICANT: FIRCA, JOSEPH R.  
CC APPLICANT: STAFFORD, DOUGLAS C.  
CC TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND  
CC TITLE OF INVENTION: PREVENTION OF C. DIFFICILE DISEASE  
CC NUMBER OF SEQUENCES: 32  
CC CORRESPONDENCE ADDRESSES:  
CC ADDRESSEE: MEDLEN & CARROLL, LLP  
CC STREET: 220 MONTGOMERY STREET, SUITE 2200  
CC CITY: SAN FRANCISCO  
CC STATE: CALIFORNIA  
CC COUNTRY: UNITED STATES OF AMERICA  
CC ZIP: 94104  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: Patentin Release #1.0, Version #1.30  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/480,604A  
CC FILING DATE: 07-JUN-1995  
CC CLASSIFICATION: 424  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 08/422,711  
CC FILING DATE: 14-APR-1995  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 08/405,496  
CC FILING DATE: 16-MAR-1995  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 08/329,154  
CC FILING DATE: 25-OCT-1994  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 08/161,907  
CC FILING DATE: 02-DEC-1993  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 07/985,321  
CC FILING DATE: 04-DEC-1992  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 07/429,791  
CC FILING DATE: 31-OCT-1989  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: INGOLIA, DIANE E.  
CC REGISTRATION NUMBER: 40,027  
CC REFERENCE/DOCKET NUMBER: OPHD-01763  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: (415) 705-8410  
CC TELEFAX: (415) 397-8338  
CC INFORMATION FOR SEQ ID NO: 8:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 91 amino acids  
CC TYPE: amino acid  
CC STRANDEDNESS: unknown  
CC TOPOLOGY: unknown  
CC MOLECULE TYPE: protein  
CC SEQUENCE 91 AA: 10432 MW: 39683 CN:  
SQ  
Query Match 38.1%; Score 59; DB 1; Length 91;  
Best Local Similarity 46.7%; Pred. No. 3.36e+01;  
Matches 7; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Db 3 KIINGKHFFYFNNDGV 17  
: ||| :|| :||  
QY 4 RTINGQLTYFRANGV 18

RESULT 11  
ID US-08-405-496A-8 STANDARD; PRT; 91 AA.  
AC xxxxxx  
XX  
DT  
XX

DE Sequence 8, Application US/08405496A  
XX Sequence 8, Application US/08405496A  
CC Patent No. 5919665  
CC GENERAL INFORMATION:  
CC APPLICANT: WILLIAMS, JAMES A.  
CC TITLE OF INVENTION: VACCINE FOR CLOSTRIDIUM BOTULINUM  
CC TITLE OF INVENTION: NEUROTOXIN  
CC NUMBER OF SEQUENCES: 30  
CC CORRESPONDENCE ADDRESSES:  
CC ADDRESSEE: MEDLEN & CARROLL, LLP  
CC STREET: 220 MONTGOMERY STREET, SUITE 2200  
CC CITY: SAN FRANCISCO  
CC STATE: CALIFORNIA  
CC COUNTRY: USA  
CC ZIP: 94104  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: Patentin Release #1.0, Version #1.30  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/405,496A  
CC FILING DATE: 16-MAR-1995  
CC CLASSIFICATION: 424  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 08/329,154  
CC FILING DATE: 25-OCT-1994  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 08/161,907  
CC FILING DATE: 02-DEC-1993  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 07/985,321  
CC FILING DATE: 04-DEC-1992  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 07/429,791  
CC FILING DATE: 31-OCT-1989  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: INGOLIA, DIANE E.  
CC REGISTRATION NUMBER: 40,027  
CC REFERENCE/DOCKET NUMBER: OPHD-01308  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: (415) 705-8410  
CC TELEFAX: (415) 397-8338  
CC INFORMATION FOR SEQ ID NO: 8:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 91 amino acids  
CC TYPE: amino acid  
CC STRANDEDNESS: unknown  
CC TOPOLOGY: unknown  
CC MOLECULE TYPE: protein  
CC SEQUENCE 91 AA: 10432 MW: 39683 CN:  
SQ  
Query Match 38.1%; Score 59; DB 2; Length 91;  
Best Local Similarity 46.7%; Pred. No. 3.36e+01;  
Matches 7; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Db 3 KIINGKHFFYFNNDGV 17  
: ||| :|| :||  
QY 4 RTINGQLTYFRANGV 18

RESULT 12  
ID US-08-485-607-6 STANDARD; PRT; 453 AA.  
AC xxxxxx  
XX  
DT  
XX  
DE Sequence 6, Application US/08485607  
XX Sequence 6, Application US/08485607  
CC Patent No. 5792627



CC REGISTRATION NUMBER: 33,268  
 CC REFERENCE/DOCKET NUMBER: ARCD:058  
 CC TELECOMMUNICATION INFORMATION:  
 CC TELEPHONE: 1-312-744-0090  
 CC TELEFAX: 1-312-755-4489  
 CC INFORMATION FOR SEQ ID NO: 6:  
 CC SEQUENCE CHARACTERISTICS:  
 CC LENGTH: 453 amino acids  
 CC TYPE: Amino acid  
 CC STRANDEDNESS: Single  
 CC TOPOLOGY: Linear  
 CC MOLECULE TYPE: Peptide  
 CC SEQUENCE 453 AA; 49608 MW; 983857 CN;

Query Match 38.1%; Score 59; DB 2; Length 453;  
 Best Local Similarity 36.8%; Pred. No. 3.36e+01;  
 Matches 7; Conservative 4; Mismatches 8; Indels 0; Gaps 0;  
 315 RIAOGEALRFRQADIDLRG 333  
 QY 4 RTINGOLLYFRANGOVYKG 22

RESULT 15  
 ID US-07-956-700B-6 STANDARD; PRT: 453 AA.  
 AC xxxxxx  
 XX  
 XX  
 DT  
 XX  
 DE

Sequence 6, Application US/07956700B

CC Sequence 6, Application US/07956700B  
 CC Patent No. 5539092  
 CC GENERAL INFORMATION:  
 CC APPLICANT: Robert Haselkorn and Piotr Gornicki  
 CC TITLE OF INVENTION: Cyanobacterial and Plant Acetyl-CoA  
 CC NUMBER OF SEQUENCES: 116  
 CC CORRESPONDENCE ADDRESS:  
 CC ADDRESSEE: Arnold, White & Durkee  
 CC STREET: 321 No. 5539092th Clark Street  
 CC CITY: Chicago  
 CC STATE: Illinois  
 CC COUNTRY: USA  
 CC ZIP: 60610  
 CC COMPUTER READABLE FORM:  
 CC MEDIUM TYPE: Floppy Disk  
 CC COMPUTER: IBM PC Compatibld  
 CC OPERATING SYSTEM: PC-DOS/MS-DOS  
 CC SOFTWARE: ASCII-DOS  
 CC CURRENT APPLICATION DATA:  
 CC APPLICATION NUMBER: US/07/956,700B  
 CC FILING DATE: 19921002  
 CC CLASSIFICATION: 800  
 CC ATTORNEY/AGENT INFORMATION:  
 CC NAME: Thomas E. No. 5539092thrup  
 CC REGISTRATION NUMBER: 33,268  
 CC REFERENCE/DOCKET NUMBER: ARCD:058  
 CC TELECOMMUNICATION INFORMATION:  
 CC TELEPHONE: 1-312-744-0090  
 CC TELEFAX: 1-312-755-4489  
 CC INFORMATION FOR SEQ ID NO: 6:  
 CC SEQUENCE CHARACTERISTICS:  
 CC LENGTH: 453 amino acids  
 CC TYPE: Amino acid  
 CC STRANDEDNESS: Single  
 CC TOPOLOGY: Linear  
 CC MOLECULE TYPE: Peptide  
 CC SEQUENCE 453 AA; 49608 MW; 983857 CN;

Query Match 38.1%; Score 59; DB 1; Length 453;  
 Best Local Similarity 36.8%; Pred. No. 3.36e+01;

Matches 7; Conservative 4; Mismatches 8; Indels 0; Gaps 0;  
 Db 315 RIAOGEALRFRQADIDLRG 333  
 QY 4 RTINGOLLYFRANGOVYKG 22

Search completed: Tue Jan 11 15:49:31 2000  
 Job time : 7 secs.

\*\*\*\*\*  
 WIRELESS (TM)  
 \*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit,  
 Copyright (c) 1993-1998 University of Edinburgh, U.K.  
 Distribution rights by Oxford Molecular Ltd

Merch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Tue Jan 11 15:41:17 2000; Maspar time 3.90 Seconds  
 109.105 Million cell updates/sec

Output not generated.

Title: >US-09-290-049-2

Description: (1-20) from US09290049.pep

Perfect Score: 135

Sequence: 1 VPYSFTRTAHDSVQDLIA 20

Scoring table: PAM 150

Gap 15

Searched: 170751 seqs, 21266608 residues

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database:

a-geneseq35  
 1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7  
 8:part8 9:part9 10:part10 11:part11 12:part12 13:part13  
 14:part14 15:part15 16:part16 17:part17 18:part18  
 19:part19 20:part20 21:part21 22:part22 23:part23  
 24:part24 25:part25 26:part26 27:part27 28:part28  
 29:part29 30:part30 31:part31 32:part32 33:part33  
 34:part34 35:part35 36:part36 37:part37 38:part38  
 39:part39

Statistics: Mean 20.274; Variance 61.611; scale 0.329

Pred. No. is the number of results predicted by chance to have a  
 score greater than or equal to the score of the result being printed,  
 and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description	Pred. No.
1	97	71.9	1592	6	R32925	Glucosyltransferase	2.44e-03
2	71	52.6	1577	16	R91047	Alpha-D-glucosyltransferase	2.33e+00
3	63	46.7	825	4	R22237	Sequence of the "g8"	1.71e+01
4	63	46.7	873	14	R79459	Infectious Laryngotracheitis	1.71e+01
5	61	45.2	344	30	W52812	Human induced tumour	2.77e+01
6	61	45.2	579	16	R86406	Human matrix metalloproteinase	2.77e+01
7	61	45.2	582	16	R86407	Human matrix metalloproteinase	2.77e+01
8	61	45.2	582	16	R75648	Human placenta derivate	2.77e+01
9	60	44.4	146	6	R32010	Rp15-tTA-1	3.53e+01
10	60	44.4	146	6	R32011	Cytotoxic T lymphocyte	3.53e+01
11	60	44.4	156	8	R40840	Glioblastoma derived	3.53e+01
12	60	44.4	375	6	R32011	TIR	3.53e+01
13	60	44.4	375	6	R32009	Rp40-tTA-1	3.53e+01
14	60	44.4	465	31	W54351	Vimentin	3.53e+01
15	59	43.7	557	24	W24096	Enzyme for production	4.48e+01
16	58	43.0	159	10	R53249	Fasciola hepatica cat	5.69e+01

17	57	42.2	1010	18	R91823	Human immunodeficient	7.21e+01
18	56	41.5	133	33	W61232	Streptococcus pneumonia	9.12e+01
19	56	41.5	1014	35	W68474	HIV-1 strain YBF30 po	1.15e+02
20	55	40.7	326	11	R57080	Fasciola hepatica cat	1.15e+02
21	55	40.7	793	11	R56502	X-prolyl-dipeptidyl a	1.15e+02
22	55	40.7	1002	37	W72993	HIV isolate LAV.MAL p	1.15e+02
23	55	40.7	1002	1	P81861	Sequence encoded by L	1.15e+02
24	55	40.7	1003	2	R08060	HIV-1 pol protein of	1.15e+02
25	54	40.0	244	21	W14444	Carb gene product.	1.45e+02
26	54	40.0	244	26	R64175	Carbapenem R.	1.45e+02
27	54	40.0	244	26	R64176	Carbapenem R.	1.45e+02
28	54	40.0	433	23	W14081	S. thermophilus exopol	1.45e+02
29	54	40.0	473	23	W22180	S. thermophilus exopol	1.45e+02
30	54	40.0	525	9	R47467	Rinder-pest virus nuc	1.45e+02
31	54	40.0	582	30	W52134	Rabbit membrane-type	1.45e+02
32	54	40.0	623	6	R31522	Carrot reverse transc	1.45e+02
33	54	40.0	775	33	W63748	Human semaphorin.	1.45e+02
34	54	40.0	845	25	W26475	KSHV glycoprotein B.	1.45e+02
35	54	40.0	1068	8	R43342	Human p110.	1.45e+02
36	54	40.0	1068	9	R46294	Ptdins 3-kinase p110	1.45e+02
37	54	40.0	1068	9	R46294	Hypo-thetical protein	1.45e+02
38	54	40.0	1076	31	W54360	Kb-binding random pep	1.83e+02
39	53	39.3	12	16	R83288	Human secreted protein	1.83e+02
40	53	39.3	285	37	W74879	Br toxin 69D1	1.83e+02
41	53	39.3	395	6	R29520	Sequence of p20, VP4,	1.83e+02
42	53	39.3	395	6	R28813	Br toxin 69D1.	1.83e+02
43	53	39.3	934	5	P20016	POI protein contained	2.30e+02
44	52	38.5	712	30	W53110	A serine/threonine pr	2.30e+02
45	52	38.5	1050	39	W67642		

# ALIGNMENTS

RESULT 1	ID	R32925	standard; Protein: 1592 AA.
AC	R32925;		
DT	28-JUN-1993 (first entry)		
DE	Glucosyltransferase I.		
KW	GT-1; Streptococcus; dental; carries.		
PN	Streptococcus sobrinus.		
PD	U05023188-A.		
PD	02-FEB-1993.		
PF	25-JUL-1991; 186592.		
PF	25-JUL-1991; JP-186592.		
PA	(FUKU/) FUKUI I.		
PA	(KARO/) KATO K.		
DR	WPI; 93-079449/10.		
DR	N-PSDB; Q37760.		
PT	DNA sequence glucosyl:transferase-I - comprises Streptococcus		
PT	sobrinus DNA sequence with at least one nucleotide added or		
PT	deleted		
PS	Claim 13; Page 15; 29pp; Japanese.		
CC	The DNA sequence from Streptococcus sobrinus strain 6715 encodes		
CC	glucosyltransferase-I (and mutants). The DNA was obt. by treating		
CC	S. sobrinus 6715 with mutanolysin, extracting the chromosomal DNA,		
CC	partially digesting with Sau3AI and fractionating on agarose gel.		
CC	The 3-5 kbp fragment was ligated into pUC18 and E. coli JM109		
CC	transformed with it. A GT-1 expressing clone was isolated and		
CC	sequenced. The clone may be used in the development of a drug for		
CC	dental caries.		
SQ	Sequence 1592 AA;		
Query Match	71.9%; Score 97; DB 6; Length 1592;		
Best Local Similarity	84.2%; Pred. No. 2.44e-03;		
Matches	16; Conservative 1; Mismatches 1; Indels 1; Gaps 1;		
Db	548 vpsysfar-ahdsveqdl1 565		
OY	1 VPYSFTRTAHDSVQDLI 19		
RESULT 2	R91047	standard; Protein: 1577 AA.	

AC R91047: 22-MAY-1996 (first entry)  
 DE Alpha-D-glucosyltransferase.  
 KW Alpha-D-glucosyltransferase: primer-independent; soluble glucan;  
 KW sucrose; transgenic plant; cloning; *Escherichia coli*;  
 KW phage lambda-Cl3; vector; plasmid pSG501; plasmid pSG502;  
 KW gene transfer; crop improvement; storage carbohydrate; pasture;  
 KW feedstuff; senescence; dextran; binder; food; pharmaceutical.  
 OS *Streptococcus salivarius* strain ATCC 25975.  
 PN M09606173-A1.  
 PD 29-FEB-1996.  
 PE 24-AUG-1995: AU00527.  
 PR 24-AUG-1994: AU-007643.  
 PA (GIEF/) GIEFARD P M.  
 PA (JACO/) JACQUES N A.  
 PA (STMP/) STIMPSON C L.  
 FI Giffard PM, Jacques NA, Simpson CL;  
 DR WP1: 96-151376/15.

Plants conty. new bacterial DNA encoding glucosyl transferase activity - retain higher levels of stored carbohydrate(s) in a form readily digestible by ruminants

PT Claim 4: Page 16-20: 31pp: English.

CC The sequence represents an alpha-D-glucosyltransferase from *Streptococcus salivarius*. The enzyme is primer-independent, and produces soluble glucan from sucrose. A gene encoding the enzyme may be cloned and expressed in *Escherichia coli* using a subclone of phage lambda-C13, e.g. plasmid pGS6501 or plasmid pGS6502. The CC DNA may also be expressed in a transgenic plant, to improve the CC level of stored carbohydrate in a pasture plant which normally contains low levels, or to prevent degradation of stored carbohydrate during plant senescence. Dextran may be isolated from the plant, for use as a food binder or pharmaceutical additive. primer independence CC ensures that the enzyme will be functional in plants. The glucan is CC poorly degraded in plants but easily degraded by bacteria in the rumen CC of grazing livestock.

9Q Sequence 1577 AA:

Query Match	52.68;	Score 71;	DB 16;	Length 1577;
Best Local Similarity	57.98;	Pred. No. 2.33e+00;		
Matches	11;	Conservative	5;	Mismatches 2; Indels 1; Gaps 1

```
Db      660 anylfvr-ahdsevgavla 67
      ::| |::| | | | | | | ::|
QY      2 PSYSFIRTAHDSEVQDLIA 20
```

RT 3  
R22237 standard; Protein; 825 AA  
P33337.

DT 23-JUL-1992 (first entry)  
DE Sequence of the "gb" homologue of ILTV encoded by the EcoRI "u" (unique),  
      same format

**KW** Subunit vaccine; immunogen; glycoprotein; promoter  
**OS** Infectious laryngotracheitis virus.

FT	peptide	1..16	/label= signal
FT			

PD	05-MAR-1992
PF	23-AUG-1991

PA (WEBS-) WEBSTER A PTY LTD.  
PI Sheppard MG, Pridaux C, Johnson M, Fahey KJ, York JJ,

DR WPI; 92-096898/12.  
DR N-PSDB; Q22989.

PT administered by aerosol  
PS Example: Fig 11; 122pp; English.

CC The inventors claim a non-infectious subunit vaccine for use against  
 CC ILTV which comprises a glycoprotein of ILTV. They also claim DNA

CC encoding the vaccine, a synthetic polypeptide displaying the  
CC antigenicity of the 205k complex or 60k ILTV glycoprotein and a  
CC recombinant ILTV with heterologous DNA inserted into a non-essential  
CC region of the genome; and a recombinant DNA mol. comprising an ILTV  
CC promoter region operatively linked to a heterologous DNA sequence.  
CC The promoter region is the ILTV gp60 promoter. (g2205 (gpb) promoter  
CC or the ORF3 promoter region. KpnI/ORF3 is located 5' of the gp60  
CC gene. Sequence 825 AA: 50

```

Query Match      46.7%; Score 63; DB 4; Length 825;
Best Local Similarity 33.3%; Pred. No. 1.71e+01;
Matches      5; Conservative      7; Mismatches      3; Indels      0; Gaps      0;

```

```
Db 570 yltivqvnaaselev 584  
      |::| :||:::  
QY 4 YSFIRTAHDSEVQDL 18
```

RESULT 4  
ID R79459 standard; Protein; 873 AA  
AC R79459.

DT	16-JAN-1996	(first entry)
DE	Infectious Laryngotracheitis	
KW	Infectious Laryngotracheitis	

KW vaccine; immunosay; detection; identification; avipox; chicken,  
fowl; probe; antisense; gene expression.  
KW Gal14d homozygous 1

	Location/Qualifiers
FH	Key
FT	misc_difference 656..659
FT	note: "these amine

FT decoded polypeptide sequence given in the  
FT specification."  
FT  
DN nsc442831-1

PD	22-AUG-1995
PF	29-OCT-1991
DP	28-OCT-1991

PR 19-NOV-1993; US-156866  
PA (UYDE ) UNIV DELAWARE.  
PI cooler ci. Poulson DT

DR WPI; 95-302091/39.  
DR N-PSDB; Q97351.  
DE T001404 Tefest40.

PT develop prods. for detection and for vaccine(s), partic. for immunising fowl.

CC The infectious laryngotracheitis virus  
CC nucleic acids can be used as probes, the

polypeptides can be used for the production of antibodies (both used in immunoassays) and in vaccines. Recombinant avipox virus

CC chickens. 873 AA;  
SQ sequence

```

Query Match      46.7%; Score 63; DB 14; Length 873;
Best Local Similarity 33.3%; Pred. NO. 1.71e+01;
Matches      5; Conservative 7; Mismatches 3; Indels

```

```

Db      618 ytfvrqvnaaeiev 632
      1:1:1 : 11:::
QY      4  YSFIRTAHDSEVQDL 18

```

RESULT	5
ID	W52812 standard; Protein; 344 AA

DE Human induced tumour protein.  
KW Human induced tumour protein; HTP; treatment; prevention; cancer.

OS	Homo sapiens.	Location/Qualifiers
FH	Key	

FT Misc\_difference 284 /note= "encoded by CTR"  
 PN WO9806846-A1.  
 PD 19-FEB-1998.  
 PF 14-AUG-1997: U14378.  
 PR 16-AUG-1996: US-689974.  
 PA (INCY-) INCYTE PHARM INC.  
 PI Au-Young J, Hawkins PR;  
 DR WPI: 98-159538/14.  
 DR N-PSDB: V20913.  
 PT Human tumour induced protein and related nucleic acid, vectors -  
 PT transformed cells and antibodies, useful for inducing  
 PT differentiation of cancer cells, and for diagnosis and monitoring of  
 PT treatment  
 PS Claim 1; Fig 1; 61pp; English.  
 CC The present sequence is human induced tumour protein (HITP),  
 CC which can be used to treat and prevent cancer, specifically  
 CC melanoma and carcinoma of the breast, colon or brain. HITP induces  
 CC cancer cell differentiation, and stops cell division.  
 SQ Sequence 344 AA;  
 DB 221 ysisirtnsqvynk1 235  
 QY 4 YSFIRTAHSEVODL 18  
 RESULT 6 45.2%; Score 61; DB 30; Length 344;  
 ID R86406 standard; Protein: 579 AA.  
 AC R86406;  
 DT 15-MAY-1996 (first entry)  
 DE Human matrix metalloprotease MMP1a.  
 KM Human; matrix metalloprotease; MMP; extracellular matrix; inhibitor;  
 KW activator; zinc-binding region.  
 OS Homo sapiens.  
 FH Key Location/Qualifiers  
 FT region 88..94  
 FT /note= "conserved region within pro-sequence,  
 FT contains a Cys residue"  
 FT region 105..110  
 FT /note= "conserved region within pro-sequence"  
 FT domain 236..246  
 FT /label= catalytic  
 FT /note= "contains 3 His residues"  
 FT region 536..556  
 FT /label= hydrophobic  
 FT WO9525171-A2.  
 DT 21-SEP-1995.  
 PR 17-MAR-1995: D00357.  
 PR 17-MAR-1994; DE-409663.  
 PR 21-OCT-1994; DE-438838.  
 PA (DELB-) DELBRUECK CENT MOLEKULARE MEDIZIN MAX.  
 PI Hinzmann B, Will H;  
 DR WPI: 95-336975/43.  
 DR N-PSDB: T03436.  
 PT New matrix metallo:protease(s) and DNA encoding them - also vectors,  
 PT recombinant cells and complexes with ligands, useful as proteolytic  
 PT agents and for identification of specific inhibitors and activators  
 PS Claim 8; Page 30; 85pp; German.  
 CC The present sequence is that of human matrix metalloprotease  
 CC MMP1a. The protein has mol. wt. 65591 and comprises a signal  
 CC peptide, pro-region, conserved catalytic domain and haemopexin-  
 CC homologous sequence, all typical of known metalloproteases. MMP1a  
 CC differs from known metalloproteases in having a hydrophobic region  
 CC near the C-terminus, suggesting that the protein is membrane-bound.  
 CC MMP proteins can hydrolyse extracellular matrix proteins and are  
 CC involved in certain pathologies. The new MMP will be useful for  
 CC detecting MMP inhibitors or activators and to generate diagnostic  
 CC antibodies.  
 SQ Sequence 579 AA;

Query Match 45.2%; Score 61; DB 16; Length 579;  
 Best Local Similarity 29.4%; Pred. No. 2,77e+01;  
 Matches 5; Conservative 8; Mismatches 4; Indels 0; Gaps 0;  
 DB 160 pyatireghekqadml 176  
 QY 3 SYSFIRTAHSEVODLI 19  
 RESULT 7  
 ID R86407 standard; Protein: 582 AA.  
 AC R86407;  
 DT 15-MAY-1996 (first entry)  
 DE Human matrix metalloprotease MMP1b.  
 KM Human; matrix metalloprotease; MMP; extracellular matrix; inhibitor;  
 KW activator; zinc-binding region.  
 OS Homo sapiens.  
 FH Key Location/Qualifiers  
 FT region 91..97  
 FT /note= "conserved region within pro-sequence,  
 FT contains a Cys residue"  
 FT region 108..113  
 FT /note= "conserved region within pro-sequence"  
 FT domain 239..249  
 FT /label= catalytic  
 FT /note= "contains 3 His residues"  
 FT region 539..559  
 FT /label= hydrophobic  
 FT WO9525171-A2.  
 DT 21-SEP-1995.  
 PR 17-MAR-1995: D00357.  
 PR 17-MAR-1994; DE-409663.  
 PR 21-OCT-1994; DE-438838.  
 PA (DELB-) DELBRUECK CENT MOLEKULARE MEDIZIN MAX.  
 PI Hinzmann B, Will H;  
 DR WPI: 95-336975/43.  
 DR N-PSDB: T03437.  
 PT New matrix metallo:protease(s) and DNA encoding them - also vectors,  
 PT recombinant cells and complexes with ligands, useful as proteolytic  
 PT agents and for identification of specific inhibitors and activators  
 PS Claim 9; Page 30-31; 85pp; German.  
 CC The present sequence is that of human matrix metalloprotease  
 CC MMP1b. The protein has mol. wt. 65900 and comprises a signal  
 CC peptide, pro-region, conserved catalytic domain and haemopexin-  
 CC homologous sequence, all typical of known metalloproteases. MMP1b  
 CC differs from known metalloproteases in having a hydrophobic region  
 CC near the C-terminus, suggesting that the protein is membrane-bound.  
 CC MMP proteins can hydrolyse extracellular matrix proteins and are  
 CC involved in certain pathologies. The new MMP will be useful for  
 CC detecting MMP inhibitors or activators and to generate diagnostic  
 CC antibodies.  
 SQ Sequence 582 AA;  
 Query Match 45.2%; Score 61; DB 16; Length 582;  
 Best Local Similarity 29.4%; Pred. No. 2,77e+01;  
 Matches 5; Conservative 8; Mismatches 4; Indels 0; Gaps 0;  
 DB 163 pyatireghekqadml 179  
 QY 3 SYSFIRTAHSEVODLI 19  
 RESULT 8  
 ID R75648 standard; Protein: 582 AA.  
 AC R75648;  
 DT 31-JAN-1996 (first entry)  
 DE Human placenta derived metalloprotease.  
 KM Human placenta; metalloprotease; probe; cancer cell detection;  
 KW monoclonal antibody; cancer; treatment; diagnosis.  
 OS Homo sapiens.  
 FH Key Location/Qualifiers  
 FT region 112..1860

```

FT      W09515374-A1. /tag- a
PN      08-JUN-1995.
PD      30-NOV-1994; J02009.
PR      30-NOV-1993; JP-341061.
PA      (FUJII) FUJII YAKUHIIN KOGYO KK.
PI      Sato H, Seki M, Shinagawa A;
DR      WPI: 95-215253/28.
N-PSDB: Q92573.
PT      Metalloprotease and monoclonal antibody recognising it - for
CC      detection of cancer cells in medical diagnosis and research.
PS      Claim 1: Pages 39-43; 67pp. Japanese
CC      Q92573 encodes R75648 a human placenta derived metalloprotease. The
CC      gene can be used as a probe for the detection of cancer cells, and
CC      monoclonal antibodies specific for the metalloprotease can be used
CC      for the treatment and diagnosis of cancer.
SQ      Sequence 582 AA;

Query Match      45.28; Score 61; DB 14; Length 582;
Best Local Similarity 29.48; Pred. No. 2,77e+01;
Matches 5; Conservative 8; Mismatches 4; Indels 0; Gaps 0;

DB      163 pvaireghekqadlm 179
QY      3 SYSFIR-TAHSEVQDLIA 19

RESULT 9
ID      R32010 standard; Protein: 146 AA.
AC      R32010;
DT      27-MAY-1993 (first entry)
DE      Rpl3-T1A-1.
KW      15 kD antigen; cytolytic lymphocyte; infection; HIV; targeting.
OS      Homo sapiens.
PN      WO9301314-A.
PF      21-JAN-1993.
PR      12-JUN-1992; U05117.
PR      10-JUL-1991; US-726607.
PR      19-FEB-1992; US-843949.
PA      (DAND ) DANA FARBER CANCER INST INC.
PI      Anderson PJ, Schlossman SF, Streuli M;
DR      WPI: 93-045515/05.
N-PSDB: Q34555.
PT      DNA encoding T1A-1 antigen - used for detecting and cytotoxic
PT      killing of cytolytic lymphocyte(s) applicable for cancer
PT      treatment
PS      Disclosure: Page 37; 61pp; English.
CC      RNA from cytolytic T cell clone T4R8C1, which expresses high levels
CC      of T1A-1 antigen was used for the construction of a cDNA library in
CC      lambda gtl1. When this expression library was screened using T1A-1
CC      monoclonal antibody, several bacteriophage expressing immunoreactive
CC      fusion proteins were identified. Cross hybridisation analysis
CC      identified three independent clones contg. related cDNAs. The
CC      largest cDNA insert was used to probe the original library for
CC      isolation of full length cDNA encoding rpl3-T1A-1, a 15 kD T1A-1
CC      antigen. The antigen is active in lymphocyte-mediated cytotoxicity
CC      and can eliminate a wide variety of virus infected or transformed
CC      target cells. T1A-1 antigen may be linked chemically or
CC      recombinantly to cell targeting ligands such as growth factors,
CC      hormone or antibodies, and may be used to kill targeted cells, e.g.
CC      cancer cells. See also R32009-17.
SQ      Sequence 146 AA;

Query Match      44.48; Score 60; DB 6; Length 146;
Best Local Similarity 31.68; Pred. No. 3.53e+01;
Matches 6; Conservative 8; Mismatches 4; Indels 1; Gaps 1;

DB      10 gysvfrfnshesaahivs 28
QY      3 SYSFIR-TAHSEVQDLIA 20

RESULT 10

```

```

ID      R12539 standard; Protein: 146 AA.
AC      R12539;
DT      24-SEP-1991 (first entry)
DE      Cytotoxic T lymphocyte-specific antigen.
KW      CTL; natural killer cells; CD8; peripheral blood; T1A-1; immunogen.
PN      EP-436400-A.
PF      10-JUL-1991.
PR      31-DEC-1990; 314456.
PR      05-JAN-1990; US-460678.
PA      (DANA-) DANA FARBER CANCER.
PI      Anderson PJ, Streuli M, Schlossman SF;
DR      WPI: 91-202096/28.
N-PSDB: Q12525.
PT      Monoclonal antibody to identify cytolytic lymphocytes - reactive
PT      with a 15kD protein in cytolytic T lymphocytes and natural killer
PT      cells
PS      Claim 10: Fig 1; 11pp; English.
CC      This sequence corresponds to a 15kD immunogenic protein associated
CC      with cytoplasmic granules in cytolytic T lymphocytes and natural
CC      killer cells. The protein is found principally in a subpopulation of
CC      CD8+ T lymphocytes from peripheral blood mononuclear cells. It
CC      reacts with monoclonal antibody T1A-1 (HB 10319).
SQ      Sequence 146 AA;

Query Match      44.48; Score 60; DB 3; Length 146;
Best Local Similarity 31.68; Pred. No. 3.53e+01;
Matches 6; Conservative 8; Mismatches 4; Indels 1; Gaps 1;

DB      10 gysvfrfnshesaahivs 28
QY      3 SYSFIR-TAHSEVQDLIA 20

RESULT 11
ID      R40840 standard; Protein: 156 AA.
AC      R40840;
DT      01-MAR-1994 (first entry)
DE      Glioblastoma derived polypeptide.
KW      Glioblastoma; differentiation; proliferation; growth; glias;
KW      neuron; immunity; tumour; aplasia.
OS      Homo sapiens.
PN      Key
PT      peptide
FT      1..23
FT      /label= Signal_peptide
FT      24..156
FT      /label= Mature_protein
PN      EP-559428-A.
PF      08-SEP-1993.
PR      02-MAR-1993; 301571.
PR      03-MAR-1992; JP-081557.
PR      22-APR-1992; JP-129558.
PA      (ONO ) ONO PHARM CO LTD.
PI      Konishi M, Miyamoto T, Naitoh T;
DR      WPI: 93-282254/36.
N-PSDB: Q47756.
PT      New polypeptide from glioblastoma cell line - useful in treatment
PT      of abnormal glial cells or neurons or for treating tumours
PS      Claims 1-4: Page 14-15; 21pp; English.
CC      The glioblastoma derived polypeptide is synthesised in and secreted
CC      from a glioblastoma cell line containing this DNA fragment and
CC      therefore may possess biological activities relating to
CC      differentiation, proliferation and growth of glias or neurons,
CC      relating to the function of immunity and relating to proliferation
CC      and growth of tumours. The polypeptide may be useful for the
CC      prevention of or in the treatment of aplasia or abnormal
CC      proliferation of glias or neurons, depression or enhancement of
CC      immunological activity. The DNA was amplified by PCR techniques
CC      using primers (Q47754-55).
SQ      Sequence 156 AA;

Query Match      44.48; Score 60; DB 8; Length 156;
Best Local Similarity 42.98; Pred. No. 3.53e+01;
Matches 9; Conservative 6; Mismatches 5; Indels 1; Gaps 1;

```



PS Disclosure; Page 35; 61pp; English

DE Enzyme for production of catechol or pyrogallol compounds.  
 KM Catalysis; enzyme; methyl group; vanillic acid; syringa; production  
 KM catechol; pyrogallol; Pseudomonas paucimobilis.

KM Catalysis; enzyme; methyl group; vanillic acid; syringa; production.  
KM catechol; pyrogallol; *Pseudomonas paucimobillis*.  
OS *Pseudomonas paucimobillis* SYK-6.

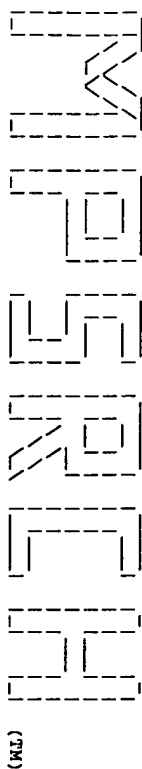
PN J09173074-A.  
PD 08-JUL-1997.  
PF 21-DEC-1995; 349914.  
PR 21-DEC-1995; JP-349914.  
PA (MAZN ) COSMO OIL CO LTD.  
PA (COSM-) COSMO SOGO KENKYUSHO KK.  
DR MPI; 97-397032/37.  
DR N-PSDB; T83645.  
PT Gene encoding enzyme which catalyses release of methyl group from  
PT vanillic or syringic acid - useful for production of catechol or  
PT pyrogallol containing molecules  
PS Claim 2: Page 8-9; 14pp; Japanese.  
CC This is an enzyme which is capable of catalysing the reaction of  
CC releasing a methyl group from vanillic acid or syringic acid. The DNA and  
CC transformed host cells can be used to produce a recombinant enzyme. This  
CC enzyme plays an important role in transforming natural components into  
CC those of catechol or pyrogallol structure.  
SO Sequence 557 AA;

Very Match 43.78; Score 59; DB 24; Length 557;  
Best Local Similarity 36.88; Pred. No. 4.48e+01;  
Matches 7; Conservative 6; Mismatches 6; Indels 0; Gaps 0;

Db 539 vpsadffikldggqiglf 557

QY 1 VPSYSFIRTAHDSVODLI 19

Search completed: Tue Jan 11 15:41:38 2000  
Job time : 21 secs.



Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

\*\*\*\*\*  
Msrch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Tue Jan 11 15:46:03 2000; Maspar time 1.64 Seconds  
145,493 Million cell updates/sec  
\*\*\*\*\*

\*\*\*\*\*  
Description: >US-09-290-049-2  
(1-20) from US09290049.pep  
Perfect Score: 135  
Sequence: 1 VPYSFRTAHSEVODLIA 20  
\*\*\*\*\*

Scoring table: PAM 150  
Gap 15

Searched: 122461 seqs, 11912985 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: a-issued  
1:5A\_COMB 2:5B\_COMB 3:PCT9\_COMB 4:backfiles1

Statistics: Mean 19.113; Variance 59.268; scale 0.322

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	61	45.2	344	2	US-08-689-Sequence 1, Applicatio	1.38e+01
2	57	42.2	160	2	US-08-726-Sequence 183, Applicat	3.61e+01
3	57	42.2	194	2	US-08-689-Sequence 3, Applicatio	3.61e+01
4	56	41.5	366	2	US-08-690-Sequence 96, Applicatio	4.58e+01
5	55	40.7	256	2	US-08-469-Sequence 58, Applicati	5.80e+01
6	55	40.7	266	2	US-07-857-Sequence 75, Applicatio	5.80e+01
7	55	40.7	282	2	US-08-701-Sequence 19, Applicatio	5.80e+01
8	55	40.7	1003	2	US-07-743-Sequence 9, Applicatio	5.80e+01
9	54	40.0	244	2	US-08-553-Sequence 2, Applicatio	7.33e+01
10	54	40.0	244	2	US-08-737-Sequence 2, Applicatio	7.33e+01
11	54	40.0	473	1	US-08-746-Sequence 13, Applicati	7.33e+01
12	54	40.0	473	1	US-08-597-Sequence 13, Applicati	7.33e+01
13	54	40.0	1069	2	US-08-162-Sequence 37, Applicati	7.33e+01
14	54	40.0	1069	2	US-08-780-Sequence 37, Applicati	7.33e+01
15	54	40.0	1080	2	US-08-780-Sequence 36, Applicati	7.33e+01
16	54	40.0	1080	2	US-08-162-Sequence 36, Applicati	7.33e+01
17	53	39.3	980	1	US-08-413-Sequence 5, Applicatio	9.24e+01
18	53	39.3	980	1	US-08-220-Sequence 5, Applicatio	9.24e+01
19	53	39.3	1172	2	US-08-313-Sequence 19, Applicati	9.24e+01
20	52	38.5	418	3	PCT-US94-0-Sequence 72, Applicati	1.16e+02
21	52	38.5	865	1	US-07-803-Sequence 13, Applicati	1.16e+02
22	52	38.5	913	2	US-07-743-Sequence 22, Applicati	1.16e+02
23	52	38.5	913	2	US-07-743-Sequence 6, Applicatio	1.16e+02

24	52	38.5	1003	2	US-07-743-Sequence 10, Applicati	1.16e+02
25	52	38.5	1003	2	US-07-743-Sequence 8, Applicatio	1.16e+02
26	52	38.5	1004	2	US-07-743-Sequence 7, Applicatio	1.16e+02
27	52	38.5	1005	2	US-07-743-Sequence 1, Applicatio	1.16e+02
28	52	38.5	1016	2	US-07-743-Sequence 2, Applicatio	1.16e+02
29	52	38.5	1016	2	US-07-743-Sequence 3, Applicatio	1.16e+02
30	52	38.5	1016	2	US-07-743-Sequence 4, Applicatio	1.16e+02
31	51	37.8	221	2	US-08-916-Sequence 9, Applicatio	1.46e+02
32	51	37.8	221	2	US-08-773-Sequence 7, Applicatio	1.46e+02
33	51	37.8	375	1	US-08-027-Sequence 2, Applicatio	1.46e+02
34	51	37.8	375	1	US-08-027-Sequence 1, Applicatio	1.46e+02
35	51	37.8	393	2	US-08-689-Sequence 4, Applicatio	1.46e+02
36	51	37.8	445	2	US-08-630-Sequence 6, Applicatio	1.46e+02
37	51	37.8	445	2	US-08-838-Sequence 4, Applicatio	1.46e+02
38	51	37.8	455	1	US-08-349-Sequence 6, Applicatio	1.46e+02
39	51	37.8	456	3	PCT-US95-1-Sequence 4, Applicatio	1.46e+02
40	51	37.8	598	2	US-08-837-Sequence 4, Applicatio	1.46e+02
41	51	37.8	718	1	US-08-445-Sequence 2, Applicatio	1.46e+02
42	51	37.8	788	2	US-07-728-Sequence 4, Applicatio	1.46e+02
43	51	37.8	1082	1	US-08-106-Sequence 32, Applicati	1.46e+02
44	51	37.8	1139	2	US-08-832-Sequence 2, Applicatio	1.46e+02
45	51	37.8	1139	2	US-08-832-Sequence 2, Applicatio	1.46e+02

## ALIGNMENTS

RESULT 1  
ID US-08-689-974-1 STANDARD; PRT: 344 AA.

Sequence 1, Application US/08689974

Sequence 1, Application US/08689974  
Patent No. 5776732

GENERAL INFORMATION:

APPLICANT: Au-Young, Janice

APPLICANT: Hawkins, Phillip R.

APPLICANT: Murray, Lynn E.

TITLE OF INVENTION: NOVEL HUMAN INDUCED TUMOR PROTEIN

NUMBER OF SEQUENCES: 5

CORRESPONDENCE ADDRESS:

ADDRESSEE: Incyte Pharmaceuticals, Inc.

STREET: 3174 Porter Drive

CITY: Palo Alto

STATE: CA

COUNTRY: U.S.

ZIP: 94304

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FASTSEQ Version 1.5

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/689,974

FILING DATE: Filed Herewith

ATTORNEY/AGENT INFORMATION:

NAME: Billings, Lucy J.

REGISTRATION NUMBER: 36,749

REFERENCE/DOCKET NUMBER: PF-0113 US

TELECOMMUNICATION INFORMATION:

TELEPHONE: 415-855-0555

TELEFAX: 415-845-4166

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 344 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: peptide

IMMEDIATE SOURCE:

```

CC LIBRARY: BRAINOT03
CC CLONE: 530522
SQ SEQUENCE 344 AA; 38258 MW; 685411 CN;

Query Match 45.2%; Score 61; DB 2; Length 344;
Best Local Similarity 53.3%; Pred. No. 1.38e+01;
Matches 8; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

Db 221 YSSTTSNNSQVNNK 235
|||:::|:|
QY 4 YSFTRTADSEVODL 18

RESULT 2 STANDARD: PRT: 160 AA.
XX US-08-726-306A-183
AC xxxxxx

DE Sequence 183, Application US/08726306A
XX Sequence 183, Application US/08726306A
CC Patent No. 5958684
CC GENERAL INFORMATION:
CC APPLICANT: van Leeuwen, Frederik Willem
CC APPLICANT: Burdach, Johannes Peter Henri
CC APPLICANT: Grosveld, Franklin G.
CC TITLE OF INVENTION: DIAGNOSIS METHOD AND REAGENTS
CC NUMBER OF SEQUENCES: 189
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Banner & Witcoff, Ltd.
CC STREET: 1 Financial Center
CC CITY: Boston
CC STATE: MA
CC COUNTRY: US
CC ZIP: 02111
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Diskette, 3.50 inch, 1.44 Mb storage
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Mordperfect 6.1
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/726,306A
CC FILING DATE: 02-Oct-1996
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: GB 95/20080.4
CC FILING DATE: 02-Oct-1995
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 60/009,832
CC FILING DATE: 01-Jan-1996
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Williams, Ph.D., Kathleen M.
CC REGISTRATION NUMBER: 34,380
CC REFERENCE/DOCKET NUMBER: 96,048-A (3255/00784)
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (617) 345-9100
CC INFORMATION FOR SEQ ID NO: 183:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 160 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: unknown
CC MOLECULE TYPE: peptide
CC CC
CC CC
CC CC
SQ SEQUENCE 160 AA; 17935 MW; 109819 CN;

Query Match 42.2%; Score 57; DB 2; Length 160;
Best Local Similarity 35.7%; Pred. No. 3.61e+01;
Matches 5; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

7 SPLKRVHEETAEI 20
||::|:|:|

```

OY		5	SFIRTAHDSSEVODL 18	
RESULT		3		
ID	US-08-689-974-3	STANDARD;	PRT;	494 AA.
XX	xxxxxx			
XX				
DT				
DE	Sequence 3, Application US/08689974			
CC	Sequence 3, Application US/08689974			
CC	Patent No. 5776732			
CC	GENERAL INFORMATION:			
CC	APPLICANT: Au-Yang, Janice			
CC	APPLICANT: Hawkins, Phillip R.			
CC	APPLICANT: Murray, Lynn E.			
CC	TITLE OF INVENTION: NOVEL HUMAN INDUCED TUMOR PROTEIN			
CC	NUMBER OF SEQUENCES: 5			
CC	CORRESPONDENCE ADDRESS:			
CC	ADDRESSEE: Incyte Pharmaceuticals, Inc.			
CC	STREET: 3174 Porter Drive			
CC	CITY: Palo Alto			
CC	STATE: CA			
CC	COUNTRY: U.S.			
CC	ZIP: 94304			
CC	COMPUTER READABLE FORM:			
CC	MEDIUM TYPE: Diskette			
CC	COMPUTER: IBM Compatible			
CC	OPERATING SYSTEM: DOS			
CC	SOFTWARE: FastSeq Version 1.5			
CC	CURRENT APPLICATION DATA:			
CC	APPLICATION NUMBER: US/08/689,974			
CC	FILING DATE: Filed Herewith			
CC	ATTORNEY/AGENT INFORMATION:			
CC	NAME: Billings, Lucy J.			
CC	REGISTRATION NUMBER: 36,749			
CC	REFERENCE/DOCKET NUMBER: PF-0113 US			
CC	TELECOMMUNICATION INFORMATION:			
CC	TELEPHONE: 415-855-0555			
CC	TELEFAX: 415-845-4166			
CC	INFORMATION FOR SEQ ID NO: 3:			
CC	SEQUENCE CHARACTERISTICS:			
CC	LENGTH: 494 amino acids			
CC	TYPE: amino acid			
CC	STRANDEDNESS: single			
CC	TOPOLOGY: linear			
CC	MOLECULE TYPE: peptide			
CC	IMMEDIATE SOURCE:			
CC	LIBRARY: GenBank			
CC	CLONE: 1293563			
SO	SEQUENCE 494 AA; 54746 MW; 1410529 CN;			
DB	350 YSIRTSTNSOVDKL 364			
OY	4 YSFIRTAHDSSEVODL 18			
RESULT	4			
ID	US-08-690-734A-96	STANDARD;	PRT;	366 AA.
XX	xxxxxx			
XX				
DT				
DE	Sequence 96, Application US/08690734A			
XX	Sequence 96, Application US/08690734A			

```

CC Patent No. 5871920
CC GENERAL INFORMATION:
CC APPLICANT: Page, David C.
CC APPLICANT: Reijo, Renee
CC TITLE OF INVENTION: DAZ: A GENE ASSOCIATED WITH AZOOSPERMIA
CC NUMBER OF SEQUENCES: 96
CC CORRESPONDENCE ADDRESSES:
CC ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
CC STREET: Two Militia Drive
CC CITY: Lexington
CC STATE: Massachusetts
CC COUNTRY: US
CC ZIP: 02173
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.30
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/690,734A
CC FILING DATE: 31-JUL-1996
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 08/310,429
CC FILING DATE: 22-SEP-1994
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Granahan, Patricia
CC REGISTRATION NUMBER: 32,227
CC REFERENCE/DOCKET NUMBER: WH194-07A
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (617) 861-6240
CC TELEFAX: (617) 861-9540
CC INFORMATION FOR SEQ ID NO: 96:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 366 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: Single
CC TOPOLOGY: Linear
CC MOLECULE TYPE: peptide
SQ SEQUENCE 366 AA; 41236 MW; 802697 CN;

Query Match 41.5%; Score 56; DB 2; Length 366;
Best Local Similarity 27.8%; Pred. No. 4.56e+01;
Matches 5; Conservative 8; Mismatches 5; Indels 0; Gaps 0;

Db 81 GYGFSEYNDVDYOKING 98
OY 3 SYSFIRTAHDSYVDLIA 20
::|: | | | |::
::|: | | | |::

F 5
T US-08-469-537A-58 STANDARD; PRT; 256 AA.
XX
AC xxxxxx
AD
DE
DT
DY
E
F
G
H
I
J
K
L
M
N
O
P
Q
R
S
T
U
V
W
X
Y
Z
aa
ac
ad
ae
af
ag
ah
ai
aj
ak
al
am
an
ao
ap
aq
ar
as
at
au
av
aw
ax
ay
az
ba
bb
bc
bd
be
bf
bg
bh
bi
bj
bk
bl
bm
bn
bo
bp
bq
br
bs
bt
bu
bv
bw
bx
by
bz
ca
cb
cc
cd
ce
cf
cg
ch
ci
cj
ck
cl
cm
cn
co
cp
cq
cr
cs
ct
cu
cv
cw
cx
cy
cz
da
db
dc
dd
de
df
dg
dh
di
dj
dk
dl
dm
dn
do
dp
dq
dr
ds
dt
du
dv
dw
dx
dy
dz
ea
eb
ec
ed
ee
ef
eg
eh
ei
ej
ek
el
em
en
eo
ep
eq
er
es
et
eu
ev
ew
ex
ey
ez
fa
fb
fc
fd
fe
ff
fg
fh
fi
fj
fk
fl
fm
fn
fo
fp
fq
fr
fs
ft
fu
fv
fw
fx
fy
fz
ga
gb
gc
gd
ge
gf
gg
gh
gi
gj
gk
gl
gm
gn
go
gp
gq
gr
gs
gt
gu
gv
gw
gx
gy
gz
ha
hb
hc
hd
he
hf
hg
hh
hi
hj
hk
hl
hm
hn
ho
hp
hq
hr
hs
ht
hu
hv
hw
hx
hy
hz
ia
ib
ic
id
ie
if
ig
ih
ii
ij
ik
il
im
in
io
ip
iq
ir
is
it
iu
iv
iw
ix
iy
iz
ja
jb
jc
jd
je
jf
jg
jh
ji
jj
jk
jl
jm
jn
jo
jp
jq
jr
js
jt
ju
jv
jw
jx
jy
jz
ka
kb
kc
kd
ke
kf
kg
kh
ki
kj
kk
kl
km
kn
ko
kp
kq
kr
ks
kt
ku
kv
kw
kx
ky
kz
la
lb
lc
ld
le
lf
lg
lh
li
lj
lk
ll
lm
ln
lo
lp
lq
lr
ls
lt
lu
lv
lw
lx
ly
lz
ma
mb
mc
md
me
mf
mg
mh
mi
mj
mk
ml
mm
mn
mo
mp
mq
mr
ms
mt
mu
mv
mw
mx
my
mz
na
nb
nc
nd
ne
nf
ng
nh
ni
nj
nk
nl
nm
nn
no
np
nq
nr
ns
nt
nu
nv
nw
nx
ny
nz
oa
ob
oc
od
oe
of
og
oh
oi
oj
ok
ol
om
on
oo
op
oq
or
os
ot
ou
ov
ow
ox
oy
oz
pa
pb
pc
pd
pe
pf
pg
ph
pi
pj
pk
pl
pm
pn
po
pp
pq
pr
ps
pt
pu
pv
pw
px
py
pz
qa
qb
qc
qd
qe
qf
qg
qh
qi
qj
qk
ql
qm
qn
qo
qp
qq
qr
qs
qt
qu
qv
qw
qx
qy
qz
ra
rb
rc
rd
re
rf
rg
rh
ri
rj
rk
rl
rm
rn
ro
rp
rq
rr
rs
rt
ru
rv
rw
rx
ry
rz
sa
sb
sc
sd
se
sf
sg
sh
si
sj
sk
sl
sm
sn
so
sp
sq
sr
ss
st
su
sv
sw
sx
sy
sz
ta
tb
tc
td
te
tf
tg
th
ti
tj
tk
tl
tm
tn
to
tp
tq
tr
ts
tt
tu
tv
tw
tx
ty
tz
ua
ub
uc
ud
ue
uf
ug
uh
ui
uj
uk
ul
um
un
uo
up
uq
ur
us
ut
uu
uv
uw
ux
uy
uz
va
vb
vc
vd
ve
vf
vg
vh
vi
vj
vk
vl
vm
vn
vo
vp
vq
vr
vs
vt
vu
vv
vw
vx
vy
vz
wa
wb
wc
wd
we
wf
wg
wh
wi
wj
wk
wl
wm
wn
wo
wp
wq
wr
ws
wt
wu
wv
ww
wx
wy
wz
xa
xb
xc
xd
xe
xf
xg
xh
xi
xj
xk
xl
xm
xn
xo
xp
xq
xr
xs
xt
xu
xv
xw
xx
xy
xz
ya
yb
yc
yd
ye
yf
yg
yh
yi
yj
yk
yl
ym
yn
yo
yp
yq
yr
ys
yt
yu
yv
yw
yx
yy
yz
za
zb
zc
zd
ze
zf
zg
zh
zi
zj
zk
zl
zm
zn
zo
zp
zq
zr
zs
zt
zu
zv
zw
zx
zy
zz

```

CC	COMPUTER READABLE FORM:
CC	MEDIUM TYPE: Diskette
CC	COMPUTER: IBM Compatible
CC	OPERATING SYSTEM: DOS
CC	SOFTWARE: FASTSEQ Version 2.0
CC	CURRENT APPLICATION DATA:
CC	APPLICATION NUMBER: US/08/469,537A
CC	FILING DATE: 06-JUN-1995
CC	CLASSIFICATION: 435
CC	PRIOR APPLICATION DATA:
CC	APPLICATION NUMBER: USSN 08/406,247
CC	FILING DATE: 17-MAR-1995
CC	APPLICATION NUMBER: USSN 08/144,992
CC	FILING DATE: 28-OCT-1993
CC	APPLICATION NUMBER: USSN 07/736,559
CC	ATTORNEY/AGENT INFORMATION:
CC	NAME: Kempler, Ph.D., Gall M
CC	REGISTRATION NUMBER: 32,143
CC	REFERENCE/DOCKET NUMBER: REG 070C
CC	TELECOMMUNICATION INFORMATION:
CC	TELEPHONE: 914-345-7400
CC	TELEFAX: 914-345-7721
CC	TELEX:
CC	INFORMATION FOR SEQ ID NO: 58:
CC	SEQUENCE CHARACTERISTICS:
CC	LENGTH: 256 amino acids
CC	TYPE: amino acid
CC	STRANDEDNESS: single
CC	TOPOLOGY: unknown
CC	MOLECULE TYPE: peptide
CC	SEQUENCE 256 AA; 29258 MW; 364977 CN;
SQ	
Db	Query Match 40.7%; Score 55; DB 2; Length 256; Best Local Similarity 47.1%; Pred.No. 5.80e+01; Matches 8; Conservative 5; Mismatches 3; Indels 1; Gaps 1.
QY	5 NFIRNETHPTVKDLG 106 ::: : :     : 5 SFIRT-AHDSEVDLLA 20
RESULT	6 STANDARD; PRT; 266 AA.
ID	US-07-857-224B-75
XX AC	xxxxxx
XX DT	
DE	Sequence 75, Application US/07857224B
XX	
CC	Sequence 75, Application US/07857224B
CC	Patent No. 5958784
CC	GENERAL INFORMATION:
CC	APPLICANT: Benner, Steven A.
CC	TITLE OF INVENTION: Predicting Folded Structures of Proteins
CC	NUMBER OF SEQUENCES: 114
CC	CORRESPONDENCE ADDRESS:
CC	ADDRESSEE: Steven A. Benner
CC	STREET: Hadlaubstrasse 151
CC	CITY: Zurich
CC	STATE: none
CC	COUNTRY: Switzerland
CC	ZIP: (note: this is an international post code) CH-8092
CC	COMPUTER READABLE FORM:
CC	MEDIUM TYPE: 3.5 inch diskette, 1.4 MB storage
CC	COMPUTER: Apple Macintosh
CC	OPERATING SYSTEM: Macintosh 7.0
CC	SOFTWARE: Microsoft Word
CC	CURRENT APPLICATION DATA:
CC	APPLICATION NUMBER: US/07/857,224B
CC	FILING DATE: 03/25/92
CC	CLASSIFICATION: 436

PRIOR APPLICATION DATA: none  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (International) 41 1 632 2830  
TELEFAX: (International) 41 1 262 2437  
TELEX: none  
INFORMATION FOR SEQ ID NO: 75:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 266  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
DESCRIPTION: protein  
ORIGINAL SOURCE:  
ORGANISM: human  
FEATURE: protein kinase; Table 8 Column 85  
PUBLICATION INFORMATION:  
AUTHORS:  
AUTHORS: Hanks, S. K.  
AUTHORS: Quinn, A. M.  
AUTHORS: Hunter, T.  
TITLE: The protein kinase family  
JOURNAL: Science  
VOLUME: 241  
PAGES: 42-52  
DATE: 1988  
SEQUENCE 266 AA; 30314 MW; 399079 CN;

Query Match 40.7%; Score 55; DB 2; Length 266;  
Best Local Similarity 47.1%; Pred. No. 5.80e+01;  
Matches 8; Conservative 5; Mismatches 3; Indels 1; Gaps 1;

Db 92 NFIRNETHNPVKDLIG 108  
:111 :1::111:  
5 SFIRT-AHDEVDLIA 20

RESULT 7  
ID US-08-701-191A-19 STANDARD; PRT; 282 AA.  
AC xxxxxx  
XX  
XX  
DT  
XX  
DE Sequence 19, Application US/08701191A  
CC Sequence 19, Application US/08701191A  
CC Patent No. 5942428  
CC GENERAL INFORMATION:  
CC APPLICANT: Moosa Mohammad, Joseph Schlessinger,  
CC APPLICANT: and Steven R. Hubbard  
CC TITLE OF INVENTION: CRYSTALS OF THE TYROSINE KINASE DOMAIN  
CC TITLE OF INVENTION: OF NON-INSULIN RECEPTOR TYROSINE KINASE  
CC NUMBER OF SEQUENCES: 41  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: Lyon & Lyon  
CC STREET: 633 West Fifth Street  
CC CITY: Los Angeles  
CC STATE: California  
CC COUNTRY: U.S.A.  
CC ZIP: 90071-2066  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: 3.5" Diskette, 1.44 MB  
CC MEDIUM TYPE: storage  
CC COMPUTER: IBM Compatible  
CC OPERATING SYSTEM: IBM P.C. DOS 5.0  
CC SOFTWARE: FASTSEQ for Windows 2.0  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08701,191A  
CC FILING DATE: August 21, 1996  
CC CLASSIFICATION: 530  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER:

FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Wardburg, Richard J.  
REGISTRATION NUMBER: 32,327  
REFERENCE/DOCKET NUMBER: 227/088  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (213) 489-1600  
TELEFAX: (213) 955-0440  
TELEX: 67-3510  
INFORMATION FOR SEQ ID NO: 19:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 282 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE 282 AA; 31925 MW; 445807 CN;

Query Match 40.7%; Score 55; DB 2; Length 282;  
Best Local Similarity 47.1%; Pred. No. 5.80e+01;  
Matches 8; Conservative 5; Mismatches 3; Indels 1; Gaps 1;

Db 103 NFIRNETHNPVKDLIG 119  
:111 :1::111:  
5 SFIRT-AHDEVDLIA 20

RESULT 8  
ID US-07-743-357-9 STANDARD; PRT; 1003 AA.  
AC xxxxxx  
XX  
XX  
DT  
XX  
DE Sequence 9, Application US/07743357  
CC Sequence 9, Application US/07743357  
CC Patent No. 5858646  
CC GENERAL INFORMATION:  
CC APPLICANT: Kang, Yong C.  
CC TITLE OF INVENTION: Polypeptide having immunological  
CC TITLE OF INVENTION: activity for use as diagnostic reagent and/or vaccine  
CC NUMBER OF SEQUENCES: 22  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: KIRBY EADES GALE BAKER  
CC STREET: Box 3432, Station D  
CC CITY: Ottawa  
CC STATE: Ontario  
CC COUNTRY: Canada  
CC ZIP: K1M 1H8  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: Patentln Release #1.0, Version #1.30  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/07743,357  
CC FILING DATE: 21-AUG-1991  
CC CLASSIFICATION: 424  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: PCT/CA90/00062  
CC FILING DATE: 23-FEB-1990  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Gale, Edwin J.  
CC REGISTRATION NUMBER: 28,584  
CC REFERENCE/DOCKET NUMBER: 30924-2  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: (613) 237-6900  
CC TELEFAX: (613) 237-0045  
CC INFORMATION FOR SEQ ID NO: 9:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 1003 amino acids  
CC TYPE: amino acid

CC STRANDEDNESS: not relevant  
CC TOPOLOGY: not relevant  
CC MOLECULE TYPE: protein  
CC HYPOTHETICAL: NO  
CC FRAGMENT TYPE: Internal  
CC ORIGINAL SOURCE:  
CC ORGANISM: Human immunodeficiency virus type 1  
CC STRAIN: MAL  
SQ SEQUENCE 1003 AA; 113673 MW; 5160362 CN;  
  
Query Match 40.7%; Score 55; DB 2; Length 1003;  
Best Local Similarity 40.0%; Pred. No. 5.80e+01;  
Matches 6; Conservative 6; Mismatches 3; Indels 0; Gaps 0;  
  
Db 508 YARIKSAHTNDYKQL 522  
1:|::|::|::|  
4 YSFIRTAHDSEVODL 18  
  
RESULT 9  
ID US-08-553-633A-2 STANDARD; PRT; 244 AA.  
xxxxxx  
  
Sequence 2, Application US/08553633A  
Patent No. 5821077  
GENERAL INFORMATION:  
APPLICANT: SALMOND, GEORGE PC  
APPLICANT: HOLDEN, MATTHEW TG  
APPLICANT: COX, ANTHONY RJ  
APPLICANT: THOMSON, NICHOLAS R  
APPLICANT: MCGOWAN, SIMON J  
TITLE OF INVENTION: PROCESS FOR ACTIVATING GENE EXPRESSION  
TITLE OF INVENTION: IN BACTERIA  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BIRCH, STEWART, KOLASCH AND BIRCH  
STREET: PO BOX 747  
CITY: FALLS CHURCH  
STATE: VA  
COUNTRY: USA  
ZIP: 22040-0747  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/553.633A  
FILING DATE: 01-APR-1996  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: SVENSSON, LEONARD R  
REGISTRATION NUMBER: 30,330  
REFERENCE/DOCKET NUMBER: 1009-104  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703)-205-8000  
TELEFAX: (703)-205-8050  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 244 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
ORIGINAL SOURCE:  
ORGANISM: Etwinia carotovora  
SQ SEQUENCE 244 AA; 28137 MW; 305391 CN;  
  
Query Match 40.0%; Score 54; DB 2; Length 244;

Best Local Similarity 35.3%; Pred. No. 7.33e+01;  
Matches 6; Conservative 6; Mismatches 5; Indels 0; Gaps 0;  
  
Db 98 SGTFTVLHDHNNVATL 114  
::|::|::|::|  
2 PSTSFIRTAHDSEVODL 18  
  
RESULT 10  
ID US-08-737-825-2 STANDARD; PRT; 244 AA.  
xxxxxx  
  
Sequence 2, Application US/08737825  
Patent No. 5871922  
GENERAL INFORMATION:  
APPLICANT: SALMOND, GEORGE PEACOCK COPELAND  
APPLICANT: MCGOWAN, SIMON JAMES  
APPLICANT: SEBATHIA, MOHAMED  
APPLICANT: COX, ANTHONY RICHARD JOHN  
APPLICANT: HOLDEN, MATTHEW THOMAS GEOFFREY  
APPLICANT: PORTER, LAUREN ELIZABETH  
APPLICANT: BYCROFT, BARBARIE WALSHAM  
APPLICANT: WILLIAMS, PAUL  
APPLICANT: STEWART, GORDON SIDNEY ANDERSON BIRNIE  
TITLE OF INVENTION: GENES INVOLVED IN THE BIOSYNTHETIC PATHWAY OF CARBAPENEM  
NUMBER OF SEQUENCES: 12  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Birch, Stewart, Kolasch & Birch, LLP  
STREET: P.O. Box 747  
CITY: Falls Church  
STATE: Virginia  
COUNTRY: USA  
ZIP: 22040-0747  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: MICROSOFT WORD97  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/737.825  
FILING DATE: 03-JUN-1997  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: SVENSSON, LEONARD R.  
REGISTRATION NUMBER: 30,330  
REFERENCE/DOCKET NUMBER: 1009-0105P  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 205-8000  
TELEFAX: (703) 205-8050  
TELEX: 248345  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 244 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Etwinia carotovora  
SQ SEQUENCE 244 AA; 28137 MW; 305391 CN;  
  
Query Match 40.0%; Score 54; DB 2; Length 244;  
Best Local Similarity 35.3%; Pred. No. 7.33e+01;  
Matches 6; Conservative 6; Mismatches 5; Indels 0; Gaps 0;  
  
Db 98 SGTFTVLHDHNNVATL 114

OY 2 PSYSFIRTAHSEVODL 18

RESULT 11  
ID US-08-746-682A-13 STANDARD; PRT: 473 AA.

Sequence 13, Application US/08746682A

Sequence 13, Application US/08746682A  
Patent No. 5786184

GENERAL INFORMATION:

APPLICANT: STINGELE, Francesca

APPLICANT: MOLLER, Beat

TITLE OF INVENTION: LACTIC BACTERIA PRODUCING

TITLE OF INVENTION: EXOPOLYSACCHARIDES

NUMBER OF SEQUENCES: 19

CORRESPONDENCE ADDRESS:

ADDRESSEE: Pennie & Edmonds

STREET: 1155 Avenue of the Americans

CITY: New York

STATE: New York

COUNTRY: U.S.A.

ZIP: 10036

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/746,682A

FILING DATE: 14-NOV-1996

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/597,236

FILING DATE: 20-JUN-1995

APPLICATION NUMBER: EP 95201669.9

FILING DATE: 20-JUN-1995

ATTORNEY/AGENT INFORMATION:

NAME: Fanucci A., Allan

REGISTRATION NUMBER: 30256

INFORMATION FOR SEQ ID NO: 13:

SEQUENCE CHARACTERISTICS:

LENGTH: 473 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

SEQUENCE 473 AA; 53028 MW; 1298864 CN;

Query Match 40.0%; Score 54; DB 2; Length 473;

Best Local Similarity 33.3%; Pred. No. 7.33e+01;

Matches 6; Conservative 6; Mismatches 6; Indels 0; Gaps 0;

DB 29 VPLYTAMTPOEYGMADL 46

1 VPSYSFIRTAHSEVODL 18

RESULT 12  
ID US-08-597-236-13 STANDARD; PRT: 473 AA.

Sequence 13, Application US/08597236

Patent No. 5733765

Sequence 13, Application US/08597236

Patent No. 5733765

Sequence 13, Application US/08597236

Patent No. 5733765

Sequence 13, Application US/08597236

Patent No. 5733765

Sequence 13, Application US/08597236

Patent No. 5733765

Sequence 13, Application US/08597236

Patent No. 5733765

GENERAL INFORMATION:

APPLICANT: STINGELE, Francesca

APPLICANT: MOLLER, Beat

TITLE OF INVENTION: LACTIC BACTERIA PRODUCING

TITLE OF INVENTION: EXOPOLYSACCHARIDES

NUMBER OF SEQUENCES: 19

CORRESPONDENCE ADDRESS:

ADDRESSEE: Pennie & Edmonds

STREET: 1155 Avenue of the Americans

CITY: New York

STATE: New York

COUNTRY: U.S.A.

ZIP: 10036

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/597,236

FILING DATE:

CLASSIFICATION: 426

PRIOR APPLICATION DATA:

APPLICATION NUMBER: EP 95201669.9

FILING DATE: 20-JUN-1995

ATTORNEY/AGENT INFORMATION:

NAME: Fanucci A., Allan

REGISTRATION NUMBER: 30256

INFORMATION FOR SEQ ID NO: 13:

SEQUENCE CHARACTERISTICS:

LENGTH: 473 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

SEQUENCE 473 AA; 53028 MW; 1298864 CN;

Query Match 40.0%; Score 54; DB 1; Length 473;

Best Local Similarity 33.3%; Pred. No. 7.33e+01;

Matches 6; Conservative 6; Mismatches 6; Indels 0; Gaps 0;

DB 29 VPLYTAMTPOEYGMADL 46

1 VPSYSFIRTAHSEVODL 18

RESULT 13  
ID US-08-162-081B-37 STANDARD; PRT: 1069 AA.

Sequence 37, Application US/08162081B

Patent No. 5824492

GENERAL INFORMATION:

APPLICANT: Hiles, Ian Donald; Fry, Michael John; Dhand, Ritu

APPLICANT: Bala; Waterfield, Michael Derek; Parker, Peter

APPLICANT: Joseph; Otsu, Masayuki; Panayotou, George; Volinia,

APPLICANT: Stefano; Gout, Ivan Tarasovitch

TITLE OF INVENTION: POLYPEPTIDES HAVING KINASE ACTIVITY,

TITLE OF INVENTION: THEIR PREPARATION AND USE

NUMBER OF SEQUENCES: 50

CORRESPONDENCE ADDRESS:

ADDRESSEE: Felte & Lynch

STREET: 805 Third Avenue

CITY: New York

STATE: New York

COUNTRY: USA

ZIP: 10022

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage



CC COMPUTER: IBM PS/2  
CC OPERATING SYSTEM: PC-DOS  
CC SOFTWARE: Wordperfect  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/162.081B  
CC FILING DATE: February 7, 1994  
CC CLASSIFICATION: 435  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: PCT/GB93/00761  
CC FILING DATE: 13 April 1993  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Pasqualini, Patricia A.  
CC REGISTRATION NUMBER: 34,894  
CC REFERENCE/DOCKET NUMBER: LUD 5256  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: (212) 688-9200  
CC TELEFAX: (212) 688-3884  
CC INFORMATION FOR SEQ ID NO: 37:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 1069 amino acids  
CC TYPE: amino acid  
CC STRANDEDNESS: single  
CC TOPOLOGY: linear  
CC SEQUENCE 1069 AA: 124437 MW: 5707845 CN:  
SQ  
Query Match 40.0%; Score 54; DB 2; Length 1069;  
Best Local Similarity 22.2%; Pred. No. 7.33e+01;  
Matches 4; Conservative 10; Mismatches 4; Indels 0; Gaps 0;  
DB 66 SSYIFSVTQEAEREFF 83  
QY 2 PSYFIRTAHDEVDPLI 19  
RESULT 14  
ID US-08-780-872-37 STANDARD; PRT: 1069 AA.  
XX xxxxxx  
AC  
XX  
DT  
XX  
DE Sequence 37, Application US/08780872  
XX  
CC Sequence 37, Application US/08780872  
CC Patent No. 5846824  
CC GENERAL INFORMATION:  
CC APPLICANT: Hiles, Ian Donald; Fry, Michael John; Dhand, Ritu  
CC APPLICANT: Bala; Waterfield, Michael Derek; Parker, Peter  
CC APPLICANT: Joseph; Otsu, Masayuki; Panayotou, George; Volinia,  
CC APPLICANT: Stefano; Gout, Ivan Tarasovitch  
CC TITLE OF INVENTION: POLYPEPTIDES HAVING KINASE ACTIVITY,  
CC NUMBER OF SEQUENCES: 50  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: Felife & Lynch  
CC STREET: 805 Third Avenue  
CC CITY: New York  
CC STATE: New York  
CC COUNTRY: USA  
CC ZIP: 10022  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage  
CC COMPUTER: IBM PS/2  
CC OPERATING SYSTEM: PC-DOS  
CC SOFTWARE: Wordperfect  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/780.872  
CC FILING DATE: 09-JAN-1997  
CC CLASSIFICATION: 435  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: 08/162.081  
CC FILING DATE: February 7, 1994  
CC APPLICATION NUMBER: PCT/GB93/00761

CC FILING DATE: 13 April 1993  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Pasqualini, Patricia A.  
CC REGISTRATION NUMBER: 34,894  
CC REFERENCE/DOCKET NUMBER: LUD 5256  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: (212) 688-9200  
CC TELEFAX: (212) 688-3884  
CC INFORMATION FOR SEQ ID NO: 37:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 1069 amino acids  
CC TYPE: amino acid  
CC STRANDEDNESS: single  
CC TOPOLOGY: linear  
CC SEQUENCE 1069 AA: 124437 MW: 5707845 CN:  
SQ  
Query Match 40.0%; Score 54; DB 2; Length 1069;  
Best Local Similarity 22.2%; Pred. No. 7.33e+01;  
Matches 4; Conservative 10; Mismatches 4; Indels 0; Gaps 0;  
DB 66 SSYIFSVTQEAEREFF 83  
QY 2 PSYFIRTAHDEVDPLI 19  
RESULT 15  
ID US-08-780-872-36 STANDARD; PRT: 1080 AA.  
XX xxxxxx  
AC  
XX  
DT  
XX  
DE Sequence 36, Application US/08780872  
XX  
CC Sequence 36, Application US/08780872  
CC Patent No. 5846824  
CC GENERAL INFORMATION:  
CC APPLICANT: Hiles, Ian Donald; Fry, Michael John; Dhand, Ritu  
CC APPLICANT: Bala; Waterfield, Michael Derek; Parker, Peter  
CC APPLICANT: Joseph; Otsu, Masayuki; Panayotou, George; Volinia,  
CC APPLICANT: Stefano; Gout, Ivan Tarasovitch  
CC TITLE OF INVENTION: POLYPEPTIDES HAVING KINASE ACTIVITY,  
CC NUMBER OF SEQUENCES: 50  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: Felife & Lynch  
CC STREET: 805 Third Avenue  
CC CITY: New York  
CC STATE: New York  
CC COUNTRY: USA  
CC ZIP: 10022  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage  
CC COMPUTER: IBM PS/2  
CC OPERATING SYSTEM: PC-DOS  
CC SOFTWARE: Wordperfect  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/780.872  
CC FILING DATE: 09-JAN-1997  
CC CLASSIFICATION: 435  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: 08/162.081  
CC FILING DATE: February 7, 1994  
CC APPLICATION NUMBER: PCT/GB93/00761  
CC FILING DATE: 13 April 1993  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Pasqualini, Patricia A.  
CC REGISTRATION NUMBER: 34,894  
CC REFERENCE/DOCKET NUMBER: LUD 5256  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: (212) 688-9200  
CC TELEFAX: (212) 688-3884  
CC INFORMATION FOR SEQ ID NO: 36:

CC SEQUENCE CHARACTERISTICS:  
 CC LENGTH: 1080 amino acids  
 CC TYPE: amino acid  
 CC STRANDEDNESS: single  
 CC TOPOLOGY: linear  
 SQ SEQUENCE 1080 AA; 125733 MW; 5831251 CN;

Query Match 40.0%; Score 54; DB 2; Length 1080;  
 Best Local Similarity 22.2%; Pred. No. 7.33e+01;  
 Matches 4; Conservative 10; Mismatches 4; Indels 0; Gaps 0;  
 Db 66 SSYIFVSVTOAEREERE 83  
 QY 2 PSISFIRAHDSVQDLI 19

Search completed: Tue Jan 11 15:46:11 2000  
 time : 8 secs.



Release 3.1A John F. Collins, Biocomputing Research Unit,  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

Run on: Tue Jan 11 15:41:54 2000; Maspar time 3.73 Seconds  
ar output not generated.

Title: >US-09-290-049-2  
Description: (1-20) from US09290049.pep  
Perfect Score: 135  
Sequence: 1 VPSYSFIRTAHDESEVDLIA 20

Scoring table: PAM 150  
Gap 15

Searched: 122810 segs, 40068593 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: p160  
1: p1r1 2: p1r2 3: p1r3 4: p1r4

Statistics: Mean 28.511; Variance 41.538; scale 0.686

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
1	115	85.2	1475	2	B3135	gtfB protein precursor	4.19e-11
2	111	82.2	1375	2	JT0345	dextranucrase (EC 2. 3.42e-07	
3	97	71.9	1592	2	A38175	glucosyltransferase P	4.36e-07
4	92	68.1	1290	2	JC5473	dextranucrase (EC 2. 3.42e-07	
5	86	63.7	1365	2	A41483	glucosyltransferase (EC 2. 3.42e-07	
6	79	58.5	1431	2	A45866	dextranucrase (EC 2. 3.42e-07	
7	78	57.8	1518	2	A44811	intermediate filament	6.06e-03
8	77	57.0	1518	2	A44811	glucosyltransferase (EC 2. 3.42e-07	
9	74	54.8	1599	2	S55395	neurofilament protein	2.32e-02
10	74	54.8	1599	2	S55395	glucosyltransferase (EC 2. 3.42e-07	
11	72	53.3	101	2	S51349	hypothetical protein	5.58e-02
12	71	52.6	532	1	QFPGM	neurofilament triplet	8.62e-02
13	71	52.6	845	2	A45669	neurofilament triplet	8.62e-02
14	71	52.6	916	2	A27864	neurofilament triplet	8.62e-02
15	70	51.9	494	2	I52658	neurofilament-66 - hu	1.33e-01
16	70	51.9	504	2	I53868	alpha-interneuron - mo	1.33e-01
17	68	50.4	505	2	A41023	hypothetical protein - ra	1.33e-01
18	68	50.4	784	2	S28296	hypothetical protein - ra	1.33e-01
19	67	49.6	798	2	PN0009	neurofilament triplet	3.11e-01
20	67	49.6	849	2	I50479	neurofilament medium	4.73e-01
21	67	49.6	849	2	S00030	neurofilament triplet	4.73e-01
22	67	49.6	858	2	S15762	neurofilament triplet	4.73e-01
23	65	48.1	472	2	S41720	intermediate filament	1.08e+00

24	65	48.1	1110	2	I51116	NF-180 - sea lamprey	1.08e+00
25	64	47.4	342	2	C48435	cysteine proteinase A	1.63e+00
26	63	46.7	771	2	E70701	probable bioF2 protei	2.44e+00
27	63	46.7	873	2	S26690	glycoprotein B - infe	2.44e+00
28	63	46.7	883	1	VGBEIL	glycoprotein B precu	2.44e+00
29	63	46.7	883	1	VGBEIL	glycoprotein B precu	2.44e+00
30	62	45.9	132	2	S66478	cytochrome P450 (CYP4	3.64e+00
31	62	45.9	287	2	JC5530	T-cluster binding pro	3.64e+00
32	62	45.9	375	2	B70184	hypothetical protein	3.64e+00
33	62	45.9	375	2	A46174	RNA-binding protein T	3.64e+00
34	62	45.9	386	2	S72435	RNA-binding protein T	3.64e+00
35	62	45.9	392	2	S72436	RNA-binding protein T	3.64e+00
36	61	45.2	290	2	E71256	probable P26 - syphil	5.41e+00
37	61	45.2	430	2	I56572	glial fibrillary acid	5.41e+00
38	61	45.2	432	2	A32936	glial fibrillary acid	5.41e+00
39	61	45.2	448	2	I48128	vimentin - Chinese ha	5.41e+00
40	61	45.2	452	2	S53906	YMK1 protein - yeast	5.41e+00
41	61	45.2	466	2	S22119	vimentin - rat	5.41e+00
42	61	45.2	466	2	A43803	vimentin - mouse	5.41e+00
43	61	45.2	582	2	I48673	matrix metalloprotein	5.41e+00
44	61	45.2	582	2	I38028	matrix metalloprotein	5.41e+00
45	61	45.2	582	2	I84471	matrix metalloprotein	5.41e+00

ALIGNMENTS

RESULT 1  
ENTRY B3135 #type complete  
TITLE gtfB protein precursor - Streptococcus mutans  
ORGANISM #formal\_name Streptococcus mutans  
DATE 23-Oct-1990 #sequence\_revision 23-Oct-1990 #text\_change 09-Sep-1997

ACCESSIONS  
REFERENCE B3135; A33128  
#authors Shiroza, T.; Ueda, S.; Kuramitsu, H. K.  
#journal J. Bacteriol. (1987) 169:4263-4270  
#cross-references NCBI:87308013  
#accession B3135

##status Preliminary  
##molecule\_type DNA  
##residues 1-1475 #label SH1  
##cross-references GB:MI7361; NID:G153639; PID:G153640  
REFERENCE A33128  
#authors Shiroza, T.; Ueda, S.; Kuramitsu, H. K.  
#submission submitted to the Protein Sequence Database, September 1990  
#accession A33128

##status Preliminary; not compared with conceptual translation  
##molecule\_type DNA  
##residues 1-171,173-641,'N',643-1475 #label SH2  
##experimental\_source strain GS-5

CLASSIFICATION #superfamily cpl repeat homology

FEATURE  
1096-1115 #domain cpl repeat homology #label CP1\  
1224-1243 #domain cpl repeat homology #label CP2\  
1289-1308 #domain cpl repeat homology #label CP3\  
1354-1373 #domain cpl repeat homology #label CP4\  
1419-1438 #domain cpl repeat homology #label CP5  
#length 1475 #molecular\_weight 165811 #checksum 7497

SUMMARY  
Query Match 85.2%; Score 115; DB 2; Length 1475;  
Best Local Similarity 95.0%; Pred. No. 4.19e-11;  
Matches 19; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

Db 552 VPSYSFIR-AHDESEVDLIA 570  
1 VPSYSFIRTAHDESEVDLIA 20  
RESULT 2  
ENTRY JT0345 #type complete  
TITLE dextranucrase (EC 2.4.1.5) precursor - Streptococcus mutans (strain GS-5)

ALTERNATE\_NAMES sucrose 6-glucosyltransferase  
 ORGANISM #formal\_name Streptococcus mutans  
 DATE 31-Mar-1992 #sequence\_revision 31-Mar-1992 #text\_change 22-Nov-1996  
 ACCESIONS JT0345; C33135  
 REFERENCE JT0345  
 #authors Ueda, S.; Shiroza, T.; Kuramitsu, H.K.  
 #journal Gene (1988) 69:101-109  
 #title Sequence analysis of the glfC gene from Streptococcus mutans GS-5.  
 #cross-references MUID:89137980  
 #accession JT0345  
 #molecule\_type DNA  
 #residues 1-1375 #label UED  
 #experimental\_source GS-5  
 REFERENCE A33135  
 #authors Shiroza, T.; Ueda, S.; Kuramitsu, H.K.  
 #journal J. Bacteriol. (1987) 169:4263-4270  
 #title Sequence analysis of the glfB gene from Streptococcus mutans.  
 #cross-references MUID:87308013  
 #accession C33135  
 #status Preliminary  
 #molecule\_type DNA  
 #residues 1-1349 #label SHI  
 #cross-references GB:M17361  
 GENETICS  
 #gene glfC  
 #FUNCTION catalyzes the synthesis of both water-soluble and water-insoluble glucans from glucose.  
 #CLASSIFICATION #superfamily cpl repeat homology  
 #KEYWORDS duplication; glycosyltransferase; hexosyltransferase; transferase  
 FEATURE  
 #domain signal sequence #status predicted #label SIG  
 1-34 #product glucosyltransferase #status predicted #label MAT  
 35-1375  
 1126-1145 #domain cpl repeat homology #label CP1\  
 1253-1272 #domain cpl repeat homology #label CP2\  
 1316-1337 #domain cpl repeat homology #label CP3  
 SUMMARY #length 1375 #molecular-weight 153021 #checksum 7015  
 Query Match 82.2%; Score 111; DB 2; Length 1375;  
 Best Local Similarity 94.7%; Pred. No. 3,42e-10;  
 Matches 18; Conservative 0; Mismatches 0; Indels 1; Gaps 1;  
 D 578 VPYSFIRAHSEVODLI 595  
 1 VPYSFIRAHSEVODLI 19  
 RESULT 3  
 ENTRY A38175 #type complete  
 TITLE glucosyltransferase precursor - Streptococcus sobrinus  
 ORGANISM #formal\_name Streptococcus sobrinus  
 DATE 28-Aug-1992 #sequence\_revision 28-Aug-1992 #text\_change 09-Sep-1997  
 ACCESIONS A38175  
 REFERENCE A38175  
 #authors Abo, H.; Msumura, T.; Kodama, T.; Ohta, H.; Fukui, K.; Kato, K.; Kagawa, H.  
 #journal J. Bacteriol. (1991) 173:989-996  
 #title Peptide sequences for sucrose splitting and glucan binding within Streptococcus sobrinus glucosyltransferase (water-insoluble glucan synthetase).  
 #cross-references MUID:91123227  
 #accession A38175  
 #status Preliminary  
 #molecule\_type DNA  
 #residues 1-1592 #label ABO  
 #cross-references GB:D90213; NID:9217032; PID:dl014946; PID:9217033  
 CLASSIFICATION #superfamily cpl repeat homology  
 FEATURE

1093-1112 #domain cpl repeat homology #label CP1\  
 1222-1241 #domain cpl repeat homology #label CP2\  
 1287-1306 #domain cpl repeat homology #label CP3\  
 1330-1351 #domain cpl repeat homology #label CP4\  
 1352-1371 #domain cpl repeat homology #label CP5\  
 1402-1420 #domain cpl repeat homology #label CP6\  
 1465-1484 #domain cpl repeat homology #label CP7\  
 1513-1532 #domain cpl repeat homology #label CP8  
 SUMMARY #length 1592 #molecular-weight 176167 #checksum 5940  
 Query Match 71.9%; Score 97; DB 2; Length 1592;  
 Best Local Similarity 84.2%; Pred. No. 4.36e-07;  
 Matches 16; Conservative 1; Mismatches 1; Indels 1; Gaps 1;  
 D 548 VPYSFIRAHSEVODLI 565  
 1 VPYSFIRAHSEVODLI 19  
 RESULT 4  
 ENTRY JC5473 #type complete  
 TITLE dextranucrase (EC 2.4.1.5) - leuconostoc mesenteroides  
 ORGANISM #formal\_name Leuconostoc mesenteroides  
 DATE 07-Jul-1997 #sequence\_revision 29-Aug-1997 #text\_change 29-Aug-1997  
 ACCESIONS JC5473  
 REFERENCE JC5473  
 #authors Monchois, V.; Willemot, R.M.; Renaud-Simeon, M.; Croux, C.; Monsan, P.  
 #journal Gene (1996) 182:23-32  
 #title Cloning and sequencing of a gene coding for a novel dextranucrase from Leuconostoc mesenteroides NRRL B-1299 synthesizing only alpha (1-6) and alpha (1-3) linkages.  
 #cross-references MUID:97136686  
 #accession JC5473  
 #status Nucleic acid sequence not shown; translation not shown  
 #molecule\_type DNA  
 #residues 1-1290 #label MON  
 #cross-references GB:U38181  
 COMMENT This enzyme catalyzes the transfer of D-glucopyranosyl units from sucrose onto acceptor molecules.  
 GENETICS  
 #gene dsra  
 #KEYWORDS glycosyltransferase; hexosyltransferase  
 FEATURE 78-870  
 922-1290  
 SUMMARY #length 1290 #molecular-weight 145572 #checksum 9386  
 Query Match 68.1%; Score 92; DB 2; Length 1290;  
 Best Local Similarity 78.9%; Pred. No. 5.15e-06;  
 Matches 15; Conservative 2; Mismatches 1; Indels 1; Gaps 1;  
 D 388 PNTSFIRAHSEVODIIA 405  
 2 PNTSFIRAHSEVODIIA 20  
 RESULT 5  
 ENTRY A41483 #type complete  
 TITLE glucosyltransferase (EC 2.4.1.-) glfs precursor - Streptococcus sobrinus  
 ORGANISM #formal\_name Streptococcus sobrinus  
 DATE 30-Jun-1992 #sequence\_revision 30-Jun-1992 #text\_change 09-Sep-1997  
 ACCESIONS A41483  
 REFERENCE A41483  
 #authors Gilmore, K.S.; Russell, R.R.B.; Ferretti, J.J.  
 #journal Infect. Immun. (1990) 58:2452-2458  
 #title Analysis of the Streptococcus downei glfs gene, which specifies a glucosyltransferase that synthesizes soluble glucans.  
 #cross-references MUID:90316665

```

##status      preliminary
##molecule_type  DNA
##residues    1-575 ##label DOD
##cross-references EMBL:X70833; NID:g312743; PID:e303308; PID:g1848062
GENETICS
#introns
CLASSIFICATION  31/1: 125/3; 220/3; 252/3; 290/3; 348/3; 445/1; 477/3; 563/3
#superfamily intermediate filament protein AV71
SUMMARY         #length 575 #molecule-weight 66359 #checksum 3801

Query Match      57.8%; Score 78; DB 2; Length 575;
Best Local Similarity 66.7%; Pred. No. 3,85e-03;
Matches 10; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Db 245 FIRRHDSFHHDLA 259
QY 6 FIRAHDSFVQDLIA 20
||| |||||:|:|
||| |||||:|:|

RESULT 8
ENTRY A44811 #type complete
TITLE glucosyltransferase (EC 2.4.1.-) I - Streptococcus salivarius
ORGANISM #formal_name Streptococcus salivarius
DATE 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change
09-Sep-1997
ACCESSIONS A44811; S22726; S28809
REFERENCE A44811
#authors Giffard, P.M.; Simpson, C.L.; Milward, C.P.; Jacques, N.A.
#journal J. Gen. Microbiol. (1991) 137:2577-2593
#title Molecular characterization of a cluster of at least two
glucosyltransferase genes in Streptococcus salivarius ATCC
25975.

#cross-references M0ID:92148377
#accession A44811
##molecule_type DNA
##residues 1-1518 ##label GIF
##cross-references EMBL:Z11873; NID:g47526; PID:g47527
##note sequence extracted from NCBI backbone (NCBIN:81050,
NCBIP:81052)

GENETICS
#gene gtfI
CLASSIFICATION #superfamily cpl repeat homology
KEYWORDS glycosyltransferase; hexosyltransferase
FEATURE
1307-1326 #domain cpl repeat homology #label CP4
SUMMARY #length 1518 #molecular-weight 167730 #checksum 1334

Query Match      57.08; Score 77; DB 2; Length 1518;
Best Local Similarity 66.7%; Pred. No. 6.06e-03;
Matches 12; Conservative 4; Mismatches 1; Indels 1; Gaps 1;

Db 604 NYVIR-AHDNNYODLIA 620
QY 3 SYSFIRAHDSFVQDLIA 20
:| ||| ||||:|:|

RESULT 9
ENTRY S55395 #type fragment
TITLE neurofilament protein M - rabbit (fragment)
ORGANISM #formal_name Oryctolagus cuniculus #common_name domestic
rabbit
DATE 15-Jul-1995 #sequence_revision 01-Sep-1995 #text_change
08-Sep-1997
ACCESSIONS S55395
REFERENCE S55395
#authors Vitalello, M.; Vettore, S.; Lamar, E.; Chien, K.R.; Gorza, L.
#submission submitted to the EMBL Data Library, January 1995
#description Neurofilament mRNA and protein are expressed in precursors of
heart conduction myocytes.

#accession S55395
#status preliminary
#molecule_type mRNA
#residues 1-644 ##label VIT
##cross-references EMBL:Z47378; NID:g854352; PID:g854353

```

```

CLASSIFICATION #superfamily cytoskeletal keratin
SUMMARY #length 644 #checksum 9233

Query Match 54.88; Score 74; DB 2; Length 644;
Best Local Similarity 56.38; Pred. No. 2,32e-02;
Matches 9; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

DB 17 AFLRTNHEEFVADLLA 32
QY 5 SFRTAHDSVEVDLLA 20
      :|:|:| |:| |:|:|

RESULT 10
ENTRY #type complete
TITLE glucosyltransferase (EC 2.4.1.-) S - Streptococcus salivarius
ORGANISM #formal_name Streptococcus salivarius
DATE 16-Sep-1992 #sequence_revision 16-Sep-1992 #text_change
       09-Sep-1997
       S22737; S28810; B44811; S22727
       S22726
AUTHORS Jacques, N.
SUBMISSION submitted to the EMBL Data Library, March 1992
ACCESSION S22737
#molecule_type DNA
#residues 1-1599 ##label JAC
#cross-references EMBL:Z11872; NID:g47530; PID:g47531
#experimental_source ATCC 25975
REFERENCE A44811
AUTHORS Giffard, P.M.; Simpson, C.L.; Milward, C.P.; Jacques, N.A.
JOURNAL J. Gen. Microbiol. (1991) 137:2577-2593
TITLE Molecular characterization of a cluster of at least two
       glucosyltransferase genes in Streptococcus salivarius ATCC
       25975.

#cross-references M01D:92148377
#accession S28810
#molecule_type DNA
#residues 1-51 ##label GIF
#cross-references EMBL:Z11873

GENETICS
#gene gltf
CLASSIFICATION #superfamily cpl repeat homology
KEYWORDS glycosyltransferase; hexosyltransferase
SUMMARY #length 1599 #molecular-weight 176480 #checksum 3135

Query Match 54.88; Score 74; DB 2; Length 1599;
Best Local Similarity 63.28; Pred. No. 2,32e-02;
Matches 12; Conservative 4; Mismatches 2; Indels 1; Gaps 1;

DB 573 ATYLPFVR-AHDSVOTVIA 590
QY 2 PSTSFIRTAHDSVEVDLLA 20
      ::|:|:| |::|:|:|

RESULT 11
ENTRY #type complete
TITLE hypothetical protein YLR346c - yeast (Saccharomyces
       cerevisiae)
ORGANISM #formal_name Saccharomyces cerevisiae
DATE 23-Feb-1995 #sequence_revision 12-May-1995 #text_change
       21-Nov-1997
ACCESSION S51349
REFERENCE S51348
AUTHORS Du, Z.
SUBMISSION submitted to the EMBL Data Library, December 1994
ACCESSION S51349
#molecule_type DNA
#residues 1-101 ##label DUZ
#cross-references EMBL:U19028; NID:g609380; PID:g609393; MIPS:YLR346c

GENETICS
#map_position 12R
SUMMARY #length 101 #molecular-weight 11348 #checksum 3891

```

```

Query Match      53.3%   Score 72:  DB 2:  Length 101;
Best Local Similarity 25.0%:  Pred. No. 5,58e-02;
Matches          5;  Conservative 12;  Mismatches 3;  Indels 0;  Gaps 0;

Db      50  ITRYDLAKAQEAEMOEHLT 69
      : 1 :::::::::::::::
Oy      1  VPYSFIRTAHDESEVODLIA 20

RESULT      12
ENTRY       QPFGM      #type fragments
TITLE       neurofilament triplet M protein - pig (fragments)
ALTERNATE_NAMES 160k neurofilament protein; NF-(medium) protein; type IV IF protein
ORGANISM    #formal_name Sus scrofa domestica #common_name domestic pig
DATE        31-Mar-1998 #sequence_revision 02-Jul-1998 #text_change 10-Jul-1998
ACCESSIONS  A05075; S02570
REFERENCE   A05075
#authors    Geisler, N.; Fischer, S.; Vandekerckhove, J.; Plessmann, U.;:
#title      Weber, K.
            EMBD J. (1984) 3:2701-2706
            Hybrid character of a large neurofilament protein (NF-M):
            intermediate filament type sequence followed by a long and
            acidic carboxy-terminal extension.
#cross-references MUID:85076594
#accession  A05075
            #molecule_type protein
            #residues 1-454 #label GEI
            #experimental_source spinal cord
REFERENCE   S02570
#authors    Geisler, N.; Vandekerckhove, J.; Weber, K.
#journal    FEBS Lett. (1987) 221:403-407
#title      Location and sequence characterization of the major
            phosphorylation sites of the high molecular mass
            neurofilament proteins M and H.
            #cross-references MUID:87304852
#accession  S02570
            #molecule_type protein
            #residues 438-450;455-459;460-475;476-514;515-532 #label GEI2
            #experimental_source spinal cord
CLASSIFICATION #superfamily cytoskeletal keratin
KEYWORDS      blocked amino end; coiled coil; intermediate filament;
            phosphoprotein

FEATURE
1-98      #domain head #status predicted #label HDV\
99-412    #domain alpha-helical rod #label ROD\
438-454,455-459, #domain tail (fragments) #status predicted #label TRIPV\
460-475,476-514 #domain tail (fragment) #status predicted #label TRP2\
515-532    #modified site blocked amino end (Ser) (probably
            acetylated) #status experimental\
456,462,465,479 #binding_site phosphate (Ser) (covalent) #status
            experimental
SUMMARY      #length 532 #checksum 5520

Query Match      52.6%:  Score 71:  DB 1:  Length 532;
Best Local Similarity 50.0%:  Pred. No. 8,62e-02;
Matches          8;  Conservative 5;  Mismatches 3;  Indels 0;  Gaps 0;

Db      232  AFLRSNHEEVADLIA 247
      ::::|::|::|::|
Oy      5  SFIRTAHDESEVODLIA 20

RESULT      13
ENTRY       A45669      #type complete
TITLE       neurofilament triplet M protein - rat
ORGANISM    #formal_name Rattus norvegicus #common_name Norway rat
DATE        31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 17-Mar-1999
ACCESSIONS  A45669; S25712; A42393

```

```

REFERENCE      A45669
#authors      Napolitano, E.W.; Chln, S.S.M.; Colman, D.R.; Liem, R.K.H.
#journal      J. Neurosci. (1987) 7:2590-2599
#title        Complete amino acid sequence and in vitro expression of rat
#             NF-M, the middle molecular weight neurofilament protein.
#cross-references MUID:87282618
#accession    A45669
#status       preliminary; not compared with conceptual translation
#molecule_type mRNA
#residues     1-845 #label NAP
#cross-references EMBL:M18628; NID:g205687; PID:g205688
REFERENCE      S25712
#authors      Kelly, B.M.; Gillespie, C.S.; Sherman, D.L.; Brophy, P.J.
#journal      J. Cell Biol. (1992) 118:397-410
#title        Schwann cells of the myelin-forming phenotype express
#             neurofilament protein NF-M.
#cross-references MUID:92332596
#accession    S25712
#status       preliminary
#molecule_type mRNA
#residues     1-17,19-21,'P',23-204,'L',206-500,'E',501-845 #label
#             KEL
#cross-references EMBL:Z12152; NID:g56751; PID:g56752
REFERENCE      A42393
#authors      Xu, Z.S.; Liu, W.S.; Willard, M.B.
#journal      J. Biol. Chem. (1992) 267:4467-4471
#title        Identification of six phosphorylation sites in the
#             COOH-terminal tail region of the rat neurofilament protein
#             M.
#cross-references MUID:92165797
#accession    A42393
#status       preliminary; not compared with conceptual translation
#molecule_type nucleic acid
#residues     411-500,'E',501-843,'D', #label XU1
#note         sequence extracted from NCBI backbone (NCBIP:83873)
CLASSIFICATION #superfamily cytoskeletal keratin
SUMMARY        #length 845 #molecular-weight 95661 #checksum 1701

Query Match      52.6%; Score 71; DB 2; Length 845;
Best Local Similarity 50.0%; Pred. No. 8.62e-02;
Matches 8; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

Db 231 AFLRSNHEEVADLLA 246
:|:|:|:|:|:|:|
QY 5 SFIRTAHDESEVDLLA 20

RESULT 14
ENTRY      A27864 #type complete
#TITLE     neurofilament triplet M protein - human
#AUTHORS   NF-M (medium) protein
#ORGANISM  #formal_name Homo sapiens #common_name man
#DATE      30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change
#           20-Mar-1998
ACCESSIONS A27864; A30157
REFERENCE   A27864
#authors    Myers, M.W.; Lazzarini, R.A.; Lee, V.M.Y.; Schlaepfer, W.W.;
#           Nelson, D.L.
#journal    EMBO J. (1987) 6:1617-1626
#title      The human mid-size neurofilament subunit: a repeated protein
#           sequence and the relationship of its gene to the
#           intermediate filament gene family.
#cross-references MUID:87275853
#accession  A27864
#molecule_type DNA
#residues   1-916 #label MYE
#cross-references GB:Y00067; NID:g35045; PID:g35046
REFERENCE   A30157
#authors    Lee, V.M.Y.; Otvos Jr., L.; Carden, M.J.; Hollosi, M.;
#           Dietzschold, B.; Lazzarini, R.A.
#journal    Proc. Natl. Acad. Sci. U.S.A. (1988) 85:1998-2002
#title      Identification of the major multiphosphorylation site in
#           mammalian neurofilaments.

```

```

#cross-references MUID:88158120
#contents      annotation; phosphorylation sites
GENETICS       360/3; 402/2
#introns
CLASSIFICATION #superfamily cytoskeletal keratin
KEYWORDS       coiled coil; phosphoprotein
FEATURE        615,628,641,654,
#binding_site phosphate (Ser) (covalent) #status
#predicted
SUMMARY        #length 916 #molecular-weight 102447 #checksum 8997

Query Match      52.6%; Score 71; DB 2; Length 916;
Best Local Similarity 50.0%; Pred. No. 8.62e-02;
Matches 8; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

Db 232 AFLRSNHEEVADLLA 247
:|:|:|:|:|:|:|
QY 5 SFIRTAHDESEVDLLA 20

RESULT 15
ENTRY      I52658 #type complete
#TITLE     neurofilament-66 - human
#AUTHORS   #formal_name Homo sapiens #common_name man
#DATE      02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change
#           16-Feb-1997
ACCESSIONS I52658
REFERENCE   I52658
#authors    Chan, S.O.; Chln, F.C.
#journal    Brain Res. Mol. Brain Res. (1995) 29:177-184
#title      Cloning and developmental expression of human 66 kd
#           neurofilament protein.
#cross-references MUID:95287809
#accession  I52658
#status       preliminary; translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues     1-494 #label RES
#cross-references GB:S78296; NID:g994843; PID:g994844
CLASSIFICATION #superfamily cytoskeletal keratin
SUMMARY        #length 494 #molecular-weight 54907 #checksum 4489

Query Match      51.9%; Score 70; DB 2; Length 494;
Best Local Similarity 50.0%; Pred. No. 1.33e-01;
Matches 8; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

Db 221 AFYRAVHDEVAELIA 236
:|:|:|:|:|:|:|
QY 5 SFIRTAHDESEVDLLA 20

Search completed: Tue Jan 11 15:42:12 2000
Job time : 18 secs.

```





\*\*\*\*\*  
 WIRENET  
 \*\*\*\*\*  
 (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.  
 Copyright (c) 1993-1998 University of Edinburgh, U.K.  
 Distribution rights by Oxford Molecular Ltd

Msrch\_pp protein - protein database search, using Smith-Waterman algorithm  
 Run on: Tue Jan 11 15:42:29 2000; Maspar time 2.40 Seconds  
 Target output not generated. 235,764 Million cell updates/sec

Title: >US-09-290-049-2  
 Description: (1-20) from US09290049.pep  
 Perfect Score: 135  
 Sequence: 1 VPSPYSTRHNDSEVDLIA 20

Scoring table: PAM 150  
 Gap 15

Searched: 77977 seqs, 28268293 residues

Post-processing: Minimum Match 0%  
 Listing first 45 summaries

Database: swiss-prot37  
 1:swissprot

Statistics: Mean 29.305; Variance 37.765; scale 0.776

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	115	85.2	1475	1	GTFR_STRMU	9.14e-13
2	111	82.2	1375	1	GTFC_STRMU	9.54e-12
3	103	76.3	1597	1	GTFL_STRMU	9.56e-10
4	97	71.9	1592	1	GTFL_STRDO	2.79e-08
5	86	63.7	1365	1	GTFS_STRDO	1.09e-05
6	79	58.5	1430	1	GTFD_STRMU	4.10e-04
7	74	54.8	644	1	NFM_RABIT	4.97e-03
8	71	52.6	454	1	NFM_PIG	2.12e-02
9	71	52.6	845	1	NFM_RAT	2.12e-02
10	71	52.6	915	1	NFM_HUMAN	2.12e-02
11	70	51.9	504	1	AINX_MOUSE	3.42e-02
12	70	51.9	505	1	AINX_RAT	3.42e-02
13	68	50.4	372	1	YLFI_CAEEL	8.75e-02
14	67	49.6	848	1	NFM_MOUSE	1.39e-01
15	67	49.6	857	1	NFM_MOUSE	1.39e-01
16	64	47.4	465	1	VIME_CHICK	5.43e-01
17	63	46.7	873	1	VIME_BOVIN	8.48e-01
18	63	46.7	883	1	VIME_ILTV6	8.48e-01
19	63	46.7	883	1	VIME_ILTV6	8.48e-01
20	62	45.9	375	1	GLYCOPROTEIN B PRECURS	1.32e+00
21	62	45.9	386	1	GLYCOPROTEIN B PRECURS	1.32e+00
22	62	45.9	392	1	GLYCOPROTEIN B PRECURS	1.32e+00
23	61	45.2	428	1	GLIAL FIBRILLARY ACIDI	2.03e+00

RESULT ID	1	STANDARD	PRT	1475 AA
AC	GTFR_STRMU			
AD	P08987			
DT	01-NOV-1988 (REL. 09, CREATED)			
DT	01-NOV-1988 (REL. 09, LAST SEQUENCE UPDATE)			
DT	15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)			
DE	GLUCOSYLTRANSFERASE-I PRECURSOR (EC 2.4.1.5) (GTF-1) (DEXTRANSUCRASE)			
DE	(SDROSE 6-GLUCOSYLTRANSFERASE).			
GN	GTFR.			
OS	STREPTOCOCCUS MUTANS.			
OC	BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; STREPTOCOCCACEAE;			
OC	STREPTOCOCCUS.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-GS-5;			
RX	MEDLINE: 87308013.			
RA	SHIROZA T., UEDA S., KURAMITSU H.K.;			
RT	"Sequence analysis of the gtfb gene from Streptococcus mutans.";			
RT	J. BACTERIOL. 169:4263-4270(1987).			
CC	- FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS, THAT ARE THOUGHT			
CC	TO PLAY A KEY ROLE IN THE DEVELOPMENT OF THE DENTAL PLAQUE BECAUSE			
CC	OF THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDIANE THE			
CC	AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS.			
CC	- CATALYTIC ACTIVITY: SUCROSE + (1,6-ALPHA-D-GLUCOSYL) (N+1).			
CC	- D-PRUCOSE + (1,6-ALPHA-D-GLUCOSYL) (N+1).			
CC	- SUBCELLULAR LOCATION: SECRETED.			
CC	- DISEASE: DENTAL CARIES.			
CC	- GTF-1 SYNTHESIZES WATER-INSOLUBLE GLUCANS (ALPHA 1,3-LINKED			
CC	GLUCOSE AND SOME 1,6 LINKAGES), GTF-S SYNTHESIZES WATER-SOLUBLE			
CC	GLUCANS (ALPHA 1,6-GLUCOSE). GTF-S SYNTHESIZES BOTH FORMS OF			
CC	- SIMILARITY: TO OTHER GLUCOSYLTRANSFERASES AND SOME TO A GLUCAN-			
CC	BINDING PROTEIN FROM S. MUTANS.			
CC	- This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (see http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
DR	EMBL: M17361; G153640; -			
DR	PIR: B33135; B33135			
DR	PFAM: PF00208; alpha-amylase; 1.			
DR	TRANSFERASE; GLUCOSYLTRANSFERASE; SIGNAL; REPEAT; DENTAL CARIES.			
FT	SIGNAL	1	34	POTENTIAL.

FT CHAIN 35 1475 GLUCOSYLTRANSFERASE-I.  
FT DOMAIN 35 1050 CATALYTIC (APPROXIMATE).  
FT TRANSFER 1096 1475 GLUCAN-BINDING (APPROXIMATE).  
FT REPEAT 1096 1129 A REPEAT.  
FT DOMAIN 1160 1469 5 X TANDEM REPEATS.  
FT REPEAT 1160 1209 1.  
FT REPEAT 1224 1274 2.  
FT REPEAT 1289 1339 3.  
FT REPEAT 1354 1404 4.  
FT REPEAT 1419 1469 5.  
SQ SEQUENCE 1475 AA; 165812 MW; 4542C1D6 CRC32;  
Query Match 85.2%; Score 115; DB 1; Length 1475;  
Best Local Similarity 95.0%; Pred. No. 9,14e-13;  
Matches 19; Conservative 0; Mismatches 52; Indels 1; Gaps 1;  
Db 552 VPSSYFIR-AHDEVDLLA 570  
1 VPSSYFIRAHDEVDLLA 20  
RESULT 2  
ID GTFC\_STRMU STANDARD; PRT; 1375 AA.  
AC P13470; P05427;  
DT 01-NOV-1988 (REL. 09, CREATED)  
DT 01-JAN-1990 (REL. 13, LAST SEQUENCE UPDATE)  
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)  
DE GLUCOSYLTRANSFERASE-SI PRECURSOR (EC 2.4.1.5) (GTF-SI)  
DE (DEXTRANSUCRASE) (SUCROSE 6-GLUCOSYLTRANSFERASE).  
GN GTFC.  
OS STREPTOCOCCUS MUTANS.  
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; STREPTOCOCCACEAE;  
OC STREPTOCOCCUS.  
RN (1)  
RP SEQUENCE FROM N.A.  
RC STRAIN-GS-5;  
RX MEDLINE: 89137980.  
RA UEDA S., SHIROGA T., KURAMITSU H.K.;  
RT "Sequence analysis of the gltf gene from Streptococcus mutans GS-5.";  
RL GENE 69:101-109(1988).  
RN (2)  
RP SEQUENCE OF 1-349 FROM N.A.  
RC STRAIN-GS-5;  
RX MEDLINE: 87308013.  
RA SHIROGA T., UEDA S., KURAMITSU H.K.;  
RT "Sequence analysis of the gltf gene from Streptococcus mutans.";  
RL J. BACTERIOL. 169:4263-4270(1987).  
CC -1- FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS, THAT ARE THOUGHT  
TO PLAY A KEY ROLE IN THE DEVELOPMENT OF THE DENTAL PLAQUE BECAUSE  
OF THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDIATE THE  
AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS.  
CC -1- CATALYTIC ACTIVITY: SUCROSE + (1,6-ALPHA-D-GLUCOSYL) (N) -  
D-FRUCTOSE + (1,6-ALPHA-D-GLUCOSYL) (N+1).  
CC -1- SUBCELLULAR LOCATION: SECRETED.  
CC -1- DISEASE: DENTAL CARIES.  
CC -1- GTF-I SYNTHESIZES WATER-INSOLUBLE GLUCANS (ALPHA 1,3-LINKED  
GLUCOSE AND SOME 1,6 LINKAGES). GTF-S SYNTHESIZES WATER-SOLUBLE  
GLUCANS (ALPHA 1,6-GLUCOSE). GTF-SI SYNTHESIZES BOTH FORMS OF  
GLUCANS.  
CC -1- SIMILARITY: TO OTHER GLUCOSYLTRANSFERASES AND SOME TO A GLUCAN-  
BINDING PROTEIN FROM S. MUTANS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: M2054; G153643; -  
CC EMBL: M17361; G153641; -  
CC PIR: J10345; J10345.

DR PIR: C33135; C33135.  
DR PFAM: PF00128; alpha-amylase; 1.  
KW TRANSFER; GLYCOSYLTRANSFERASE; SIGNAL; REPEAT; DENTAL CARIES.  
FT SIGNAL 1 34  
FT CHAIN 35 1375 GLUCOSYLTRANSFERASE-SI.  
FT DOMAIN 35 1050 CATALYTIC (APPROXIMATE).  
FT TRANSFER 1126 1375 GLUCAN-BINDING (APPROXIMATE).  
FT DOMAIN 1126 1375 2.4 A, 1 C AND 1 AC REPEATS.  
FT REPEAT 1126 1159 A REPEAT.  
FT REPEAT 1169 1200 A REPEAT.  
FT REPEAT 1227 1238 C REPEAT.  
FT REPEAT 1253 1303 AC REPEAT.  
FT REPEAT 1318 1330 A REPEAT (INCOMPLETE).  
SQ SEQUENCE 1375 AA; 153022 MW; 3EA3727E CRC32;  
Query Match 82.2%; Score 111; DB 1; Length 1375;  
Best Local Similarity 94.7%; Pred. No. 9,54e-12;  
Matches 18; Conservative 0; Mismatches 0; Indels 1; Gaps 1;  
Db 578 VPSSYFIR-AHDEVDLLI 595  
1 VPSSYFIRAHDEVDLLI 19  
OY 1 VPSSYFIRAHDEVDLLI 19  
RESULT 3  
ID GTF1\_STRDO STANDARD; PRT; 1597 AA.  
AC P11001;  
DT 01-JUL-1989 (REL. 11, CREATED)  
DT 01-JUL-1989 (REL. 11, LAST SEQUENCE UPDATE)  
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)  
DE GLUCOSYLTRANSFERASE-1 PRECURSOR (EC 2.4.1.5) (GTF-1) (DEXTRANSUCRASE)  
DE (SUCROSE 6-GLUCOSYLTRANSFERASE).  
GN GTF1.  
OS STREPTOCOCCUS DOWNEI (STREPTOCOCCUS SOBRINUS).  
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; STREPTOCOCCACEAE;  
OC STREPTOCOCCUS.  
RN (1)  
RP SEQUENCE FROM N.A.  
RC STRAIN-MFE28;  
RX MEDLINE: 87308014.  
RA FERRETTI J.J., GILPIN M.L., RUSSELL R.R.B.;  
RT "Nucleotide sequence of a glucosyltransferase gene from Streptococcus  
sibirinus MFE28.";  
RL J. BACTERIOL. 169:4271-4278(1987).  
CC -1- FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS, THAT ARE THOUGHT  
TO PLAY A KEY ROLE IN THE DEVELOPMENT OF THE DENTAL PLAQUE BECAUSE  
OF THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDIATE THE  
AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS.  
CC -1- CATALYTIC ACTIVITY: SUCROSE + (1,6-ALPHA-D-GLUCOSYL) (N) -  
D-FRUCTOSE + (1,6-ALPHA-D-GLUCOSYL) (N+1).  
CC -1- SUBCELLULAR LOCATION: SECRETED.  
CC -1- DISEASE: DENTAL CARIES.  
CC -1- GTF-I SYNTHESIZES WATER-INSOLUBLE GLUCANS (ALPHA 1,3-LINKED  
GLUCOSE AND SOME 1,6 LINKAGES). GTF-S SYNTHESIZES WATER-SOLUBLE  
GLUCANS (ALPHA 1,6-GLUCOSE). GTF-SI SYNTHESIZES BOTH FORMS OF  
GLUCANS.  
CC -1- SIMILARITY: TO OTHER GLUCOSYLTRANSFERASES AND SOME TO A GLUCAN-  
BINDING PROTEIN FROM S. MUTANS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: M17391; G153647; -  
CC PFAM: PF00128; alpha-amylase; 1.  
KW TRANSFERASE; GLYCOSYLTRANSFERASE; SIGNAL; REPEAT; DENTAL CARIES.  
FT SIGNAL 1 38  
FT CHAIN 39 1050 POTENTIAL.  
FT DOMAIN 39 1050 GLUCOSYLTRANSFERASE-I.  
FT CATALYTIC (APPROXIMATE).

FT DOMAIN 1099 1597 GLUCAN-BINDING (APPROXIMATE).  
 FT DOMAIN 1099 1597 1.25 A, 2 B, AND 5 AC REPEATS.  
 FT REPEAT 1099 1597 A REPEAT.  
 FT REPEAT 1163 1213 AC REPEAT.  
 FT REPEAT 1227 1277 AC REPEAT.  
 FT REPEAT 1292 1342 AC REPEAT.  
 FT REPEAT 1352 1399 B REPEAT.  
 FT REPEAT 1406 1455 AC REPEAT.  
 FT REPEAT 1465 1512 B REPEAT.  
 FT REPEAT 1519 1568 AC REPEAT.  
 FT REPEAT 1582 1597 A REPEAT (INCOMPLETE).  
 SQ SEQUENCE 1597 AA: 177080 MW: 995233CA CRC32:

Query Match 76.3%; Score 103; DB 1; Length 1597;  
 Best Local Similarity 89.5%; Pred. No. 9,56e-10;  
 Matches 17; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

Db 554 VPSYSFAR-AHDSVODLI 571  
 QY 1 VPSYSFIRTAHDSVODLI 19  
 ID GTF2\_STRDO STANDARD; PRT: 1592 AA.  
 AC P27470:  
 DT 01-AUG-1992 (REL. 23, CREATED)  
 DT 01-AUG-1992 (REL. 23, LAST SEQUENCE UPDATE)  
 DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)  
 DE GLUCOSYLTRANSFERASE-1 PRECURSOR (EC 2.4.1.5) (GTF-1) (DEXTRANSUCRASE)  
 DE (SUCCROSE 6-GLUCOSYLTRANSFERASE).  
 OS STREPTOCOCCUS DOWNEI (STREPTOCOCCUS SOBRINUS).  
 OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; STREPTOCOCCACEAE;  
 OC STREPTOCOCCUS.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-6715:  
 RX MEDLINE: 91123227.  
 RA ABO H., MATSUMURA T., KODAMA T., OHTA H., FUKUI K., KATO K.,  
 RA KAGAWA H.;  
 RT "Peptide sequences for sucrose splitting and glucan binding within  
 RT streptococcus sobrinus glucosyltransferase (water-insoluble glucan  
 RT synthetase).";  
 RT J. BACTERIOL. 173:989-996(1991).  
 CC -1- FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS, THAT ARE THOUGHT  
 CC TO PLAY A KEY ROLE IN THE DEVELOPMENT OF THE DENTAL PLAQUE BECAUSE  
 CC OF THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDIATE THE  
 CC AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS.  
 CC -1- CATALYTIC ACTIVITY: SUCROSE + (1,6-ALPHA-D-GLUCOSYL)(N) -  
 CC D-FRUCTOSE + (1,6-ALPHA-D-GLUCOSYL)(N+1).  
 CC -1- SUBCELLULAR LOCATION: SECRETED.  
 CC -1- DISEASE: DENTAL CARIES.  
 CC -1- GTF-1 SYNTHESIZES WATER-INSOLUBLE GLUCANS (ALPHA 1,3-LINKED  
 CC GLUCOSE AND SOME 1,6 LINKAGES), GTF-S SYNTHESIZES WATER-SOLUBLE  
 CC GLUCANS (ALPHA 1,6-GLUCOSE). GTF-SI SYNTHESIZES BOTH FORMS OF  
 CC GLUCANS.  
 CC -1- SIMILARITY: TO OTHER GLUCOSYLTRANSFERASES AND SOME TO A GLUCAN-  
 CC BINDING PROTEIN FROM S. MUTANS.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC EMBL: D90213; G217033; -  
 DR PIR: A38175; A38175.  
 DR PFAM: PF00128; alpha-amyase; 1.  
 DR HSSP: P00695; 2HEE.  
 KW TRANSFERASE; GLUCOSYLTRANSFERASE; SIGNAL; REPEAT; DENTAL CARIES.  
 FT SIGNAL 1 38 POTENTIAL.  
 FT CHAIN 39 1592 GLUCOSYLTRANSFERASE-1.

FT DOMAIN 39 1044 CATALYTIC (APPROXIMATE).  
 FT DOMAIN 1093 1592 GLUCAN-BINDING (APPROXIMATE).  
 FT REPEAT 1093 1592 6.5 X TANDEM REPEATS.  
 FT REPEAT 1093 1442 1.  
 FT REPEAT 1158 1207 2.  
 FT REPEAT 1222 1272 3.  
 FT REPEAT 1287 1337 4.  
 FT REPEAT 1402 1451 5.  
 FT REPEAT 1514 1563 6.  
 FT REPEAT 1577 1592 7 (INCOMPLETE).  
 SQ SEQUENCE 1592 AA: 176167 MW: 273233FA CRC32:

Query Match 71.9%; Score 97; DB 1; Length 1592;  
 Best Local Similarity 84.2%; Pred. No. 2,79e-08;  
 Matches 16; Conservative 1; Mismatches 1; Indels 1; Gaps 1;

Db 548 VPSYSFAR-AHDSVODLI 565  
 QY 1 VPSYSFIRTAHDSVODLI 19  
 ID GTF5\_STRDO STANDARD; PRT: 1365 AA.  
 AC P29336:  
 DT 01-DEC-1992 (REL. 24, CREATED)  
 DT 01-DEC-1992 (REL. 24, LAST SEQUENCE UPDATE)  
 DT 01-APR-1993 (REL. 25, LAST ANNOTATION UPDATE)  
 DE GLUCOSYLTRANSFERASE-5 PRECURSOR (EC 2.4.1.5) (GTF-S) (DEXTRANSUCRASE)  
 DE (SUCCROSE 6-GLUCOSYLTRANSFERASE).  
 GN GTF5.  
 OS STREPTOCOCCUS DOWNEI (STREPTOCOCCUS SOBRINUS).  
 OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; STREPTOCOCCACEAE;  
 OC STREPTOCOCCUS.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-MEE28;  
 RX MEDLINE: 90316665.  
 RA GILMORE K.S., RUSSELL R.R., FERRETTI J.J.;  
 RT "Analysis of the Streptococcus downei gtf5 gene, which specifies a  
 RT glucosyltransferase that synthesizes soluble glucans.";  
 RT INFECT. IMMUN. 58:2452-2458(1990).  
 CC -1- FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS, THAT ARE THOUGHT  
 CC TO PLAY A KEY ROLE IN THE DEVELOPMENT OF THE DENTAL PLAQUE BECAUSE  
 CC OF THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDIATE THE  
 CC AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS.  
 CC -1- CATALYTIC ACTIVITY: SUCROSE + (1,6-ALPHA-D-GLUCOSYL)(N) -  
 CC D-FRUCTOSE + (1,6-ALPHA-D-GLUCOSYL)(N+1).  
 CC -1- ENZYME REGULATION: GLUCAN SYNTHESIS BY GTF-S IS INDEPENDENT OF  
 CC PRIMER GLUCAN UNLIKE GTF-1.  
 CC -1- DISEASE: DENTAL CARIES.  
 CC -1- GTF-S SYNTHESIZES WATER-SOLUBLE GLUCANS (ALPHA 1,6-GLUCOSE).  
 CC -1- SIMILARITY: TO OTHER GLUCOSYLTRANSFERASES AND SOME TO A GLUCAN-  
 CC BINDING PROTEIN FROM S. MUTANS.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC EMBL: M30943; G139653; -  
 DR PIR: A14483; A14483.  
 DR PFAM: PF00128; alpha-amyase; 1.  
 KW TRANSFERASE; GLUCOSYLTRANSFERASE; SIGNAL; REPEAT; DENTAL CARIES.  
 FT SIGNAL 1 36 OR 37 (POTENTIAL).  
 FT CHAIN 37 1365 GLUCOSYLTRANSFERASE-S.  
 FT DOMAIN 1083 1365 CATALYTIC (APPROXIMATE).  
 FT DOMAIN 1083 1365 GLUCAN-BINDING (APPROXIMATE).  
 FT REPEAT 1083 1131 4.5 X TANDEM REPEATS.  
 FT REPEAT 1150 1199 2.



DE NEUROFILAMENT TRIPLET M PROTEIN (160 KD NEUROFILAMENT PROTEIN) (NF-M)  
 DE (FRAGMENT).  
 GN NEFM.  
 OS SUS SCROFA (PIG).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;  
 OC ARTIODACTYLIA; SUIFORMES; SUINA; SUIDAE; SUS.  
 RP [1]  
 RP SEQUENCE.  
 RC TISSUE-SPINAL CORD;  
 RX MEDLINE: 85076594.  
 RA GEISLER N., FISCHER S., VANDEKERCKHOVE J., PLESSMANN U., WEBER K.;  
 RT "Hybrid character of a large neurofilament protein (NF-M):  
 RT intermediate filament type sequence followed by a long and acidic  
 RT carboxy-terminal extension.";  
 RL EMBO J. 3:2701-2706(1984).  
 CC -1- FUNCTION: NEUROFILAMENTS USUALLY CONTAIN THREE IF PROTEINS: L, M,  
 CC AND H WHICH ARE INVOLVED IN THE MAINTENANCE OF NEURONAL CALIBER.  
 CC -1- PTM: THERE ARE A NUMBER OF REPEATS OF THE TRIPEPTIDE K-S-P, NFM IS  
 CC PHOSPHORYLATED ON A NUMBER OF THE SERINES IN THIS MOTIF. IT IS  
 CC THOUGHT THAT PHOSPHORYLATION OF NFM RESULTS IN THE FORMATION OF  
 CC INTERFILAMENT CROSS BRIDGES THAT ARE IMPORTANT IN THE MAINTENANCE  
 CC OF AXONAL CALIBER.  
 CC -1- PTM: PHOSPHORYLATION SEEMS TO PLAY A MAJOR ROLE IN THE FUNCTIONING  
 CC OF THE LARGER NEUROFILAMENT POLYPEPTIDES (NF-M AND NF-H), THE  
 CC LEVELS OF PHOSPHORYLATION BEING ALTERED DEVELOPMENTALLY AND  
 CC COINCIDENT WITH A CHANGE IN THE NEUROFILAMENT FUNCTION.  
 CC -1- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.  
 CC PIR: A05075; OEPGM.  
 DR PROSITE: PS00226; IF: 1.  
 DR PFAM: PF00038; filament; 1.  
 DR INTERMEDIATE FILAMENT; HEPTAD REPEAT PATTERN; COILED COIL; NEURONE;  
 KW ACETYLATION; PHOSPHORYLATION; GLYCOPROTEIN.  
 KM MOD\_RES 1 1  
 FT DOMAIN 1 104  
 FT MOD\_RES 1 104  
 FT DOMAIN 105 412  
 FT MOD\_RES 105 412  
 FT DOMAIN 105 >454  
 FT MOD\_RES 105 >454  
 FT DOMAIN 137 149  
 FT MOD\_RES 137 149  
 FT DOMAIN 150 248  
 FT MOD\_RES 150 248  
 FT DOMAIN 249 265  
 FT MOD\_RES 249 265  
 FT DOMAIN 266 287  
 FT MOD\_RES 266 287  
 FT DOMAIN 288 291  
 FT MOD\_RES 288 291  
 FT DOMAIN 292 411  
 FT MOD\_RES 292 411  
 FT CARBOHYD 47 47  
 FT MOD\_RES 47 47  
 FT CARBOHYD 432 432  
 FT MOD\_RES 432 432  
 FT NON\_TER 454 454  
 FT MOD\_RES 454 454  
 SQ SEQUENCE 454 AA; 51854 MW; 3DFC58A1 CRC32;  
 Query Match 52.6%; Score 71; DB 1; Length 454;  
 Best Local Similarity 50.0%; Pred. No. 2,12e-02;  
 Matches 8; Conservative 5; Mismatches 3; Indels 0; Gaps 0;  
 Db 232 AFLRSNHEEVADLLA 247  
 QY 5 SFIRTAHSEVODLLA 20

RT the middle molecular weight neurofilament protein.";  
 RL J. NEUROSCI. 7:2590-2599(1987).  
 RN [2]  
 RP PHOSPHORYLATION SITES, AND REVISION TO 500.  
 RX MEDLINE: 92165797.  
 RA XU Z.-S., LIU W.-S., WILLARD M.B.;  
 RT "Identification of six phosphorylation sites in the COOH-terminal  
 RT tail region of the rat neurofilament protein M.";  
 RL J. BIOL. CHEM. 267:4467-4471(1992).  
 RN [3]  
 RP CARBOHYDRATE-BINDING SITES.  
 RX MEDLINE: 93346421.  
 RA DONG D.L.-Y., XU Z.-S., CHEVRIER M.R., COTTER R.J., CLEVELAND D.W.,  
 RA HART G.W.;  
 RT "Glycosylation of mammalian neurofilaments. Localization of multiple  
 RT O-linked N-acetylglucosamine moieties on neurofilament polypeptides  
 RT L and M.";  
 RL J. BIOL. CHEM. 268:16679-16687(1993).  
 CC -1- FUNCTION: NEUROFILAMENTS USUALLY CONTAIN THREE IF PROTEINS: L, M,  
 CC AND H WHICH ARE INVOLVED IN THE MAINTENANCE OF NEURONAL CALIBER.  
 CC -1- PTM: THERE ARE A NUMBER OF REPEATS OF THE TRIPEPTIDE K-S-P, NFM IS  
 CC PHOSPHORYLATED ON A NUMBER OF THE SERINES IN THIS MOTIF. IT IS  
 CC THOUGHT THAT PHOSPHORYLATION OF NFM RESULTS IN THE FORMATION OF  
 CC INTERFILAMENT CROSS BRIDGES THAT ARE IMPORTANT IN THE MAINTENANCE  
 CC OF AXONAL CALIBER.  
 CC -1- PTM: PHOSPHORYLATION SEEMS TO PLAY A MAJOR ROLE IN THE FUNCTIONING  
 CC OF THE LARGER NEUROFILAMENT POLYPEPTIDES (NF-M AND NF-H), THE  
 CC LEVELS OF PHOSPHORYLATION BEING ALTERED DEVELOPMENTALLY AND  
 CC COINCIDENT WITH A CHANGE IN THE NEUROFILAMENT FUNCTION.  
 CC -1- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL: M18628; G205688; -  
 DR PROSITE: PS00226; IF: 1.  
 DR PFAM: PF00038; filament; 1.  
 DR INTERMEDIATE FILAMENT; HEPTAD REPEAT PATTERN; COILED COIL; NEURONE;  
 KW PHOSPHORYLATION; GLYCOPROTEIN.  
 KM MOD\_RES 0 0  
 FT INIT\_MET 0 0  
 FT DOMAIN 1 103  
 FT MOD\_RES 1 103  
 FT DOMAIN 103 410  
 FT MOD\_RES 103 410  
 FT DOMAIN 103 844  
 FT MOD\_RES 103 844  
 FT DOMAIN 135 147  
 FT MOD\_RES 135 147  
 FT DOMAIN 148 246  
 FT MOD\_RES 148 246  
 FT DOMAIN 247 263  
 FT MOD\_RES 247 263  
 FT DOMAIN 264 285  
 FT MOD\_RES 264 285  
 FT DOMAIN 286 289  
 FT MOD\_RES 286 289  
 FT DOMAIN 290 410  
 FT MOD\_RES 290 410  
 FT CARBOHYD 47 47  
 FT MOD\_RES 47 47  
 FT CARBOHYD 430 430  
 FT MOD\_RES 430 430  
 FT MOD\_RES 502 502  
 FT MOD\_RES 506 506  
 FT MOD\_RES 536 536  
 FT MOD\_RES 603 603  
 FT MOD\_RES 608 608  
 FT MOD\_RES 666 666  
 FT MOD\_RES 666 666  
 FT CONFLICT 500 500  
 SQ SEQUENCE 845 AA; 95660 MW; 8C2B516C CRC32;  
 Query Match 52.6%; Score 71; DB 1; Length 845;  
 Best Local Similarity 50.0%; Pred. No. 2,12e-02;  
 Matches 8; Conservative 5; Mismatches 3; Indels 0; Gaps 0;  
 Db 230 AFLRSNHEEVADLLA 245  
 QY 5 SFIRTAHSEVODLLA 20

RA NAPOLITANO E.W., CHIN S.S.M., COLMAN D.R., LIEM R.K.H.;  
 "Complete amino acid sequence and in vitro expression of rat NF-M,"





```

CC INTERFILAMENT CROSS BRIDGES THAT ARE IMPORTANT IN THE MAINTENANCE
CC OF AXONAL CALIBER.
CC -1- PPM: PHOSPHORYLATION SEEMS TO PLAY A MAJOR ROLE IN THE FUNCTIONING
CC OF THE LARGER NEUROFILAMENT POLYPEPTIDES (NF-M AND NF-H), THE
CC LEVELS OF PHOSPHORYLATION BEING ALTERED DEVELOPMENTALLY AND
CC COINCIDENT WITH A CHANGE IN THE NEUROFILAMENT FUNCTION.
CC -1- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X05640; G297529;
DR EMBL: M20481; G300040;
DR PIR: S00030; S00030.
DR MGD: MGI:97314; NFM.
DR PROSITE: PS00226; IF; 1.
DR PFM: PFO0038; filament; 1.
CC INTERMEDIATE FILAMENT; HEPTAD REPEAT PATTERN; COILED COIL; NEURONE;
CC PHOSPHORYLATION; GLYCOPROTEIN.
CC -----
FT INT_MET 0 0
FT DOMAIN 1 102 HEAD.
FT DOMAIN 103 410 ROD.
FT DOMAIN 411 848 TAIL.
FT DOMAIN 103 134 COIL 1A.
FT DOMAIN 135 147 LINKER 1.
FT DOMAIN 148 246 COIL 1B.
FT DOMAIN 247 263 LINKER 12.
FT DOMAIN 264 285 COIL 2A.
FT DOMAIN 286 289 LINKER 2.
FT DOMAIN 290 410 COIL 2B.
FT CARBOHYD 47 47 GLCNAC (BY SIMILARITY).
FT CARBOHYD 430 430 GLCNAC (BY SIMILARITY).
FT CONFLICT 432 432 S -> F (IN REF. 2).
FT CONFLICT 539 540 QA -> RR (IN REF. 2).
SQ SEQUENCE 848 AA; 95910 MW; E06A637A CRC32;

Query Match 49.6%; Score 67; DB 1; Length 848;
Best Local Similarity 50.0%; Pred. No. 1.39e-01;
Matches 8; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

DB 230 AFLRNHEEVADLLA 245
:|:|:|:|:|:|
5 SFRTADSEVODLLA 20

RESULT 15
ID NFM_CHICK STANDARD; PRT; 857 AA.
AC P16053;
DT 01-APR-1990 (REL. 14, CREATED)
DT 01-APR-1990 (REL. 14, LAST SEQUENCE UPDATE)
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE NEUROFILAMENT TRIPLET M PROTEIN (160 KD NEUROFILAMENT PROTEIN) (NF-M).
GN NFM.
OS GALLUS GALLUS (CHICKEN).
OC EUKARYOTA; METAZOA; VERTEBRATA; ARCHOSAURIA; AVES;
OC NEOGNATHAE; GALLIFORMES; PHASIANIDAE; PHASIANINAE; GALLUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 90174973.
RA ZOPF D., DINEVA B., BETZ H., GUNDELINGER E.D.;
RT "Isolation of the chicken middle-molecular weight neurofilament
RT (NF-M) gene and characterization of its promoter.";
RL NUCLEIC ACIDS RES. 18:521-529(1990).
RN [2]
RP SEQUENCE OF 259-857 FROM N.A.
RX MEDLINE: 88112814.
RA ZOPF D., HERMANS-BORGMEYER I., GUNDELINGER E.D., BETZ H.;
RT "Identification of gene products expressed in the developing chick

```

```

RT visual system: characterization of a middle-molecular-weight
RT neurofilament cDNA.";
RL GENES DEV. 1:699-708(1987).
CC -1- FUNCTION: NEUROFILAMENTS USUALLY CONTAIN THREE IF PROTEINS: L, M,
CC AND H WHICH ARE INVOLVED IN THE MAINTENANCE OF NEURONAL CALIBER.
CC -1- PPM: THERE ARE A NUMBER OF REPEATS OF THE TRIPLETTE K-S-P, NFM IS
CC PHOSPHORYLATED ON A NUMBER OF THE SERINES IN THIS MOTIF. IT IS
CC THOUGHT THAT PHOSPHORYLATION OF NFM RESULTS IN THE FORMATION OF
CC INTERFILAMENT CROSS BRIDGES THAT ARE IMPORTANT IN THE MAINTENANCE
CC OF AXONAL CALIBER.
CC -1- PPM: PHOSPHORYLATION SEEMS TO PLAY A MAJOR ROLE IN THE FUNCTIONING
CC OF THE LARGER NEUROFILAMENT POLYPEPTIDES (NF-M AND NF-H), THE
CC LEVELS OF PHOSPHORYLATION BEING ALTERED DEVELOPMENTALLY AND
CC COINCIDENT WITH A CHANGE IN THE NEUROFILAMENT FUNCTION.
CC -1- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X17102; G63689;
DR EMBL: X05558; G63686;
DR PIR: A27040; A27040.
DR PIR: S08061; S08061.
DR PIR: S15762; S15762.
DR PROSITE: PS00226; IF; 1.
DR PFM: PFO0038; filament; 1.
CC INTERMEDIATE FILAMENT; HEPTAD REPEAT PATTERN; COILED COIL; NEURONE;
CC PHOSPHORYLATION; GLYCOPROTEIN.
CC -----
FT INT_MET 0 0
FT DOMAIN 1 98 HEAD.
FT DOMAIN 99 406 ROD.
FT DOMAIN 407 857 TAIL.
FT DOMAIN 99 130 COIL 1A.
FT DOMAIN 131 143 LINKER 1.
FT DOMAIN 144 242 COIL 1B.
FT DOMAIN 243 259 LINKER 12.
FT DOMAIN 260 281 COIL 2A.
FT DOMAIN 282 285 LINKER 2.
FT DOMAIN 286 406 COIL 2B.
FT CARBOHYD 46 46 GLCNAC (BY SIMILARITY).
FT CARBOHYD 426 426 GLCNAC (BY SIMILARITY).
FT CONFLICT 546 546 G -> R (IN REF. 2).
SQ SEQUENCE 857 AA; 95704 MW; 3D05FFDD CRC32;

Query Match 49.6%; Score 67; DB 1; Length 857;
Best Local Similarity 43.8%; Pred. No. 1.39e-01;
Matches 7; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

DB 226 AFLRNHEEVADLLA 241
:|:|:|:|:|:|
5 SFRTADSEVODLLA 20

```

Search completed: Tue Jan 11 15:42:37 2000  
Job time : 8 secs.



\*\*\*\*\*  
 WISE (TM)  
 \*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
 Copyright (c) 1993-1998 University of Edinburgh, U.K.  
 Distribution rights by Oxford Molecular Ltd

MSrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Tue Jan 11 15:36:00 2000; MasPar time 5.50 Seconds  
 81.178 Million cell updates/sec

Target output not generated.

Title: >US-09-290-049-1

Description: (1-21) from US09290049.pep

Perfect Score: 148

Sequence: 1 ANDHLSILEAMSDNDPYLHD 21

Scoring table: PAM 150  
 Gap 15

Searched: 170751 seqs, 21266608 residues

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database:

a-geneseg35  
 1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7  
 8:part8 9:part9 10:part10 11:part11 12:part12 13:part13  
 14:part14 15:part15 16:part16 17:part17 18:part18  
 19:part19 20:part20 21:part21 22:part22 23:part23  
 24:part24 25:part25 26:part26 27:part27 28:part28  
 29:part29 30:part30 31:part31 32:part32 33:part33  
 34:part34 35:part35 36:part36 37:part37 38:part38  
 39:part39

Statistics: Mean 20.949; Variance 72.779; scale 0.288

Pred. No. is the number of results predicted by chance to have a  
 score greater than or equal to the score of the result being printed,  
 and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	131	88.5	1592	6	R32925 Glucosyltransferase	5.55e-06
2	83	56.1	1577	16	R91047 Alpha-D-glucosyltransferase	6.09e-01
3	66	44.6	387	33	W62297 Scenedesmus D1 protease	2.73e+01
4	63	42.6	388	33	W62298 Wheat D1 protease	5.20e+01
5	62	41.9	648	22	W17045 Mouse c-raf 1 protein	6.43e+01
6	62	41.9	648	22	W17044 Human c-raf 1 protein	6.43e+01
7	62	41.9	648	22	W17046 Mouse mutant c-raf 1	6.43e+01
8	62	41.9	648	22	W17047 Mouse mutant c-raf 1	6.43e+01
9	62	41.9	648	4	R22560 Mouse mutant c-raf 1	6.43e+01
10	62	41.9	648	4	R22562 Mouse mutant c-raf 1	6.43e+01
11	62	41.9	648	4	R22561 Mouse mutant c-raf 1	6.43e+01
12	62	41.9	648	21	W13107 Human c-raf-1	6.43e+01
13	62	41.9	648	21	W13107 Human c-raf-1	6.43e+01
14	62	41.9	648	22	W17048 Mouse mutant c-raf 1	6.43e+01
15	62	41.9	648	22	W17049 Mouse mutant c-raf 1	6.43e+01
16	62	41.9	648	32	W62220 Raf-1 protein	6.43e+01

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
17	62	41.9	648	4	R22563 Mouse mutant c-raf-1	6.43e+01
18	62	41.9	648	18	R22559 Mouse c-raf-1	6.43e+01
19	62	41.9	648	18	R22559 Mouse c-raf-1	6.43e+01
20	60	40.5	210	3	R15613 Yeast alpha 2 mating	9.81e+01
21	60	40.5	236	38	W90021 Expressed antigen for	9.81e+01
22	60	40.5	239	38	W89935 Antigen 1 from Cluste	9.81e+01
23	60	40.5	298	34	W71486 Helicobacter polypept	9.81e+01
24	60	40.5	473	33	W62300 Tobacco D1 protease	9.81e+01
25	59	39.9	178	1	P90061 Human Fc gamma recept	1.21e+02
26	59	39.9	338	12	R65965 T. niyeum GAPDH	1.21e+02
27	59	39.9	576	17	R85891 WD-40 domain-contg. p	1.21e+02
28	58	39.9	755	39	W81365 Human prothromone conv	1.21e+02
29	58	39.9	67	31	W28316 Staphylococcus aureus	1.49e+02
30	58	39.9	158	11	R56770 Fibrin binding peptid	1.49e+02
31	58	39.9	159	28	W45492 Targeting ligand derl	1.49e+02
32	58	39.9	167	14	R78615 MPD-1 extracellular r	1.49e+02
33	58	39.9	288	9	R47801 Murine programmed cel	1.49e+02
34	58	39.9	288	14	R78614 MPD-1	1.49e+02
35	58	39.9	392	39	W67891 Human secreted protei	1.49e+02
36	58	39.9	596	14	R78616 Expression vector pME	1.49e+02
37	58	39.9	724	20	W10730 Human Factor XIIIa re	1.49e+02
38	58	39.9	731	3	R14376 Factor XIII subunit a	1.49e+02
39	58	39.9	732	5	R25385 Human Factor XIII	1.49e+02
40	58	39.9	732	2	P70293 Sequence of human fac	1.49e+02
41	58	39.9	761	1	P82920 A and A' subunits of	1.49e+02
42	58	39.9	771	15	R80495 Japanese oyster trans	1.49e+02
43	58	39.9	783	35	W64560 Candida albicans o1g	1.49e+02
44	57	38.5	222	29	W55287 H. pylori ORF 13ep120	1.83e+02
45	57	38.5	404	30	W55726 H. pylori ORF hp8e100	1.83e+02

# ALIGNMENTS

RESULT 1  
 ID R32925 standard; Protein: 1592 AA.  
 AC R32925;  
 DT 28-JUN-1993 (first entry)  
 DE Glucosyltransferase I.  
 KW GT-1; Streptococcus; dental; caries.  
 OS Streptococcus sobrinus.  
 PN J05023188.A.  
 PD 02-FEB-1993.  
 PF 25-JUL-1991; 186592.  
 PR 25-JUL-1991; JP-186592.  
 PA (FKRU/) FKRU I.  
 PA (KARO/) KARO K.  
 DR WPI: 93-079449/10.  
 DR N-PSDB: Q37760.  
 PT DNA sequence glucosyl:transferase-I - comprises Streptococcus  
 PT sobrinus DNA sequence with at least one nucleotide added or  
 PT deleted  
 PS Claim 13, Page 15; 29pp; Japanese.  
 CC The DNA sequence from Streptococcus sobrinus strain 6715 encodes  
 CC glucosyltransferase-I (and mutants). The DNA was obt. by treating  
 CC S. sobrinus 6715 with mutanolysin, extracting the chromosomal DNA,  
 CC partially digesting with Sau3AI and fractionating on an agarose gel.  
 CC The 3-5 kbp fragment was ligated into pUC18 and E. coli JM109  
 CC transformed with it. A GT-1 expressing clone was isolated and  
 CC sequenced. The clone may be used in the development of a drug for  
 CC dental caries.  
 CC sequence 1592 AA;  
 SQ  
 Query Match 88.5%; Score 131; DB 6; Length 1592;  
 Best Local Similarity 85.7%; Pred. No. 5.55e-06;  
 Matches 18; Conservative 3; Mismatches 0; Indels 0; Gaps 0;  
 DB 477 annhvsivawsdndpylhd 497  
 ||:|||||:|||||  
 QY 1 ANDHLSILEAMSDNDPYLHD 21  
 RESULT 2  
 ID R91047 standard; Protein: 1577 AA.

AC R91047;  
 DT 22-MAY-1996 (first entry)  
 DE Alpha-D-glucosyltransferase.  
 CC Alpha-D-glucosyltransferase; primer-independent; soluble glucan;  
 CC sucrose; transgenic plant; cloning; Escherichia coli;  
 CC phage lambda-C13; vector; plasmid pGSG501; plasmid pGSG502;  
 CC gene transfer; crop improvement; storage carbohydrate; pasture;  
 CC feedstuff; senescence; dextran; binder; food; pharmaceutical.  
 OS Streptococcus salivarius strain ATCC 25975.  
 PN M09606173-A1.  
 PD 29-FEB-1996.  
 PF 24-AUG-1995: AU0527.  
 PR 24-AUG-1994: AU-007643.  
 PA (GIF/) GIFFARD P M.  
 PA (JACO/) JACQUES N A.  
 PA (SIMP/) SIMPSON C L.  
 PI Giffard PM, Jacques NA, Simpson CL;  
 DP WPI: 96-151376/15.  
 N-PSDB: T13139.  
 CC Plants contg. new bacterial DNA encoding glucosyl transferase  
 PT activity - retain higher levels of stored carbohydrate(s) in a form  
 PT readily digestible by ruminants  
 PS Claim 4: Page 16-20: 31pp: English.  
 CC The sequence represents an alpha-D-glucosyltransferase from  
 CC Streptococcus salivarius. The enzyme is primer-independent, and  
 CC produces soluble glucan from sucrose. A gene encoding the enzyme  
 CC may be cloned and expressed in Escherichia coli using a subclone  
 CC of phage lambda-C13, e.g. plasmid pGSG501 or plasmid pGSG502. The  
 CC DNA may also be expressed in a transgenic plant, to improve the  
 CC level of stored carbohydrate in a pasture plant which normally  
 CC contains low levels, or to prevent degradation of stored carbohydrate  
 CC during plant senescence. Dextran may be isolated from the plant, for  
 CC use as a food binder or pharmaceutical additive. Primer independence  
 CC ensures that the enzyme will be functional in plants. The glucan is  
 CC poorly degraded in plants but easily degraded by bacteria in the rumen  
 CC of grazing livestock.  
 CC Sequence 1577 AA;  
 SQ

Query Match 56.1%; Score 83; DB 16; Length 1577;  
 Best Local Similarity 72.2%; Pred. No. 6.09e-01;  
 Matches 13; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

DB 588 alahslleawsyndhy 605  
 QY 1 ANDHLSTLEAWSDNDTPY 18

DT 28-SEP-1998 (first entry)  
 DE Scenedesmus D1 protease.  
 CC Scenedesmus; D1 protease; herbicide; inhibition; alga; wheat;  
 CC detection; identification.  
 OS Scenedesmus obliquus.  
 PN M09824934-A2.  
 PD 11-JUN-1998.  
 PF 03-DEC-1997: U21964.  
 PR 05-DEC-1996: US-759581.  
 PA (DUPO) DU PONT DE NEMOURS & CO E I.  
 PI Chisholm DA, Diner BA, Donaldson GK, Hershey HP,  
 PI Jordan DB, Tang XS, Trost JT, Wang S, Warren PV;  
 DP WPI: 98-333349/29.  
 N-PSDB: V39818.  
 CC New isolated plant D1 protease(s) - used for identifying herbicidal  
 PT agents which target the D1 protease by inhibiting the enzyme  
 PT activity.  
 PS Claim 23: Page 44-45: 73pp: English.  
 CC A method has been developed for identifying a herbicidal agent which  
 CC inhibits D1 protease. The method comprises: (a) incubating a D1 protease  
 CC enzyme in a sample suspected of containing a herbicidal agent with a  
 CC suitable D1 enzyme substrate, where an enzyme product is formed; and (b)  
 CC detecting and quantifying the enzyme product of step (a). Also described

CC In the present invention are: (1) a method for detecting in vivo a  
 CC herbicidal agent which inhibits D1 protease comprising: (a) preparing a  
 CC reaction mixture comprising: (1) a wild type cell comprising an active D1  
 CC protease enzyme capable of processing a D1 pre-protein; and a photosystem  
 CC II core complex capable of variable chlorophyll fluorescence; (11) a  
 CC suspected herbicidal agent which inhibits D1 protease; and (111) a suitable  
 CC growth medium; (b) illuminating the reaction mixture at illumination  
 CC conditions to permit D1 turnover; and (c) measuring the variable  
 CC chlorophyll fluorescence produced in step (b), where the level of the  
 CC variable chlorophyll fluorescence correlates with herbicidal activity of  
 CC the suspected herbicidal agent; (2) a gene encoding a D1 protease enzyme  
 CC isolated from: (a) wheat; (b) tobacco; (c) spinach; (d) Scenedesmus; or  
 CC (e) Synechocystis; (see V39818 to V39822, which encode W62297 to W62301  
 CC respectively). Native or recombinant D1 proteases can be used to develop  
 CC assays for the detection of herbicidal compositions capable of  
 CC inhibiting D1.  
 CC Sequence 387 AA;  
 SQ

Query Match 44.6%; Score 66; DB 33; Length 387;  
 Best Local Similarity 36.8%; Pred. No. 2.73e+01;  
 Matches 7; Conservative 7; Mismatches 5; Indels 0; Gaps 0;

DB 2 tseqllfleaavrayv 20  
 QY 1 ANDHLSTLEAWSDNDTPYL 19

RESULT 4  
 AC W62298; 388 AA.  
 DT 28-SEP-1998 (first entry)  
 DE Wheat D1 protease.  
 CC Scenedesmus; D1 protease; herbicide; inhibition; alga; wheat;  
 CC detection; identification.  
 OS Triticum sp.  
 PN M09824934-A2.  
 PD 11-JUN-1998.  
 PF 03-DEC-1997: U21964.  
 PR 05-DEC-1996: US-759581.  
 PA (DUPO) DU PONT DE NEMOURS & CO E I.  
 PI Chisholm DA, Diner BA, Donaldson GK, Hershey HP,  
 PI Jordan DB, Tang XS, Trost JT, Wang S, Warren PV;  
 DP WPI: 98-333349/29.  
 N-PSDB: V39819.  
 CC New isolated plant D1 protease(s) - used for identifying herbicidal  
 PT agents which target the D1 protease by inhibiting the enzyme  
 PT activity.  
 PS Claim 23: Page 47-48: 73pp: English.  
 CC A method has been developed for identifying a herbicidal agent which  
 CC inhibits D1 protease. The method comprises: (a) incubating a D1 protease  
 CC enzyme in a sample suspected of containing a herbicidal agent with a  
 CC suitable D1 enzyme substrate, where an enzyme product is formed; and (b)  
 CC detecting and quantifying the enzyme product of step (a). Also described  
 CC in the present invention are: (1) a method for detecting in vivo a  
 CC herbicidal agent which inhibits D1 protease comprising: (a) preparing a  
 CC reaction mixture comprising: (1) a wild type cell comprising an active D1  
 CC protease enzyme capable of processing a D1 pre-protein; and a photosystem  
 CC II core complex capable of variable chlorophyll fluorescence; (11) a  
 CC suspected herbicidal agent which inhibits D1 protease; and (111) a suitable  
 CC growth medium; (b) illuminating the reaction mixture at illumination  
 CC conditions to permit D1 turnover; and (c) measuring the variable  
 CC chlorophyll fluorescence produced in step (b), where the level of the  
 CC variable chlorophyll fluorescence correlates with herbicidal activity of  
 CC the suspected herbicidal agent; (2) a gene encoding a D1 protease enzyme  
 CC isolated from: (a) wheat; (b) tobacco; (c) spinach; (d) Scenedesmus; or  
 CC (e) Synechocystis; (see V39818 to V39822, which encode W62297 to W62301  
 CC respectively). Native or recombinant D1 proteases can be used to develop  
 CC assays for the detection of herbicidal compositions capable of  
 CC inhibiting D1.  
 CC Sequence 388 AA;  
 SQ

Query Match 42.6%; Score 63; DB 33; Length 388;  
 Best Local Similarity 38.9%; Pred. No. 5.20e+01;







```

RESULT 15
ID W17049 standard; Protein; 648 AA.
AC W17049;
DT 16-JUL-1997 (first entry)
DE Mutant mouse c-raf-1 protein used in diagnosis of cancer.
KW raf; oncogene; lymphoma; lung cancer; neoplasia; point mutation;
KM conserved region; adenocarcinoma; codon 533; diagnosis; detection.
OS Mus musculus.
PN US5618670-A.
PD 08-APR-1997.
PF 26-AUG-1988; 236947.
PR 26-AUG-1988; US-236947.
PR 16-SEP-1991; US-759738.
PR 24-JAN-1994; US-185282.
PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.
PI Rapp UR, Storm SM;
DR WPI: 97-225421/20.
PT Classification of lymphoma or lung cancer - on the basis of a point
  mutation in c-raf-1 gene.
  Example 3; Column 23-28; 26pp; English.
CC W17046-W17049 are mutant versions of the mouse c-raf-1 protein, derived
CC from four different types of tumour. They were used in a method for
CC classifying a lymphoma or lung cancer, by comparison with the wild-type
CC c-raf-1 protein sequence derived from different tumours in mice. The
CC method, when used for diagnosing human tumours, involves detecting the
CC presence of a point mutation in a conserved region of the c-raf-1 gene
CC (codon 533, encoding Ser - this mutant version has Phe at this
CC position) derived from lymphoma or lung cancer tissue and classifying
CC the lymphoma or lung cancer as a c-raf-1 mutation-associated cancer if
CC one or more point mutations are present. The method is particularly
CC applicable to diagnosis of lung adenocarcinoma.
SQ Sequence 648 AA;

```

```

Query Match 41.9%; Score 62; DB 22; Length 648;
Best Local Similarity 30.0%; Pred. No. 6.43e+01;
Matches 6; Conservative 7; Mismatches 7; Indels 0; Gaps 0;

```

```

Db 413 tkdnlavtgcgsslykh 432

```

```

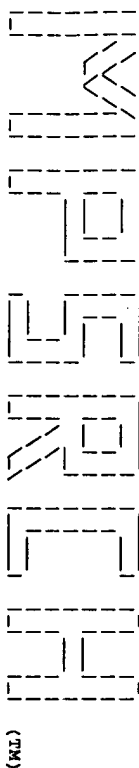
OY 1 ANDHLSILEAMSDMDTPYLH 20

```

```

Search completed: Tue Jan 11 15:36:37 2000
Job time : 37 secs.

```



Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MSrch\_PP protein - protein database search, using Smith-Waterman algorithm  
Run on: Tue Jan 11 15:38:57 2000; MasPar time 1.65 Seconds  
151.295 Million cell updates/sec  
MasPar output not generated.

Title: >US-09-290-049-1  
Description: (1-21) From US09290049.pep  
Perfect Score: 148  
Sequence: 1 ANDHLSTLEAWSDNDTPYLMD 21

Scoring table: PAM 150  
Gap 15

Searched: 122461 seqs, 11912985 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: a-issued  
1:5A\_COMB 2:5B\_COMB 3:PCT9\_COMB 4:backfill1

Statistics: Mean 19.740; Variance 68.890; scale 0.287

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Match Length	DB	ID	Description	Pred. No.
1	66	44.6	387	2	US-08-759- Sequence 4, Applicatio	1.25e+01
2	66	44.6	464	2	US-08-759- Sequence 22, Applicati	1.25e+01
3	63	42.6	388	2	US-08-759- Sequence 9, Applicatio	2.41e+01
4	62	41.9	267	2	US-07-857- Sequence 42, Applicatio	3.00e+01
5	62	41.9	315	1	US-08-909- Sequence 12, Applicati	3.00e+01
6	62	41.9	315	1	US-08-909- Sequence 12, Applicati	3.00e+01
7	62	41.9	315	1	US-08-571- Sequence 12, Applicati	3.00e+01
8	62	41.9	346	1	US-08-276- Sequence 5, Applicatio	3.00e+01
9	62	41.9	648	2	US-08-185- Sequence 12, Applicati	3.00e+01
10	62	41.9	648	2	US-08-886- Sequence 6, Applicatio	3.00e+01
11	62	41.9	648	1	US-08-276- Sequence 2, Applicatio	3.00e+01
12	62	41.9	648	1	US-08-185- Sequence 3, Applicatio	3.00e+01
13	62	41.9	648	1	US-08-185- Sequence 1, Applicatio	3.00e+01
14	62	41.9	648	1	US-08-185- Sequence 4, Applicatio	3.00e+01
15	62	41.9	648	1	US-08-185- Sequence 5, Applicatio	3.00e+01
16	62	41.9	648	1	US-08-185- Sequence 2, Applicatio	3.00e+01
17	61	41.2	480	3	PCT-US95-0 Sequence 12, Applicati	3.72e+01
18	61	41.2	480	3	US-08-272- Sequence 12, Applicati	3.72e+01
19	60	40.5	473	2	US-08-759- Sequence 15, Applicati	4.61e+01
20	59	39.9	576	1	US-08-190- Sequence 56, Applicati	5.71e+01
21	58	39.2	105	1	US-08-459- Sequence 4, Applicatio	7.06e+01
22	58	39.2	770	1	US-08-525- Sequence 1, Applicatio	7.06e+01
23	58	39.2	771	1	US-08-525- Sequence 3, Applicatio	7.06e+01

24	57	38.5	184	2	US-08-737- Sequence 10, Applicati	8.72e+01
25	57	38.5	500	4	5171684-2 Patent No. 5171684.	8.72e+01
26	56	37.8	161	2	US-08-403- Sequence 22, Applicati	1.08e+02
27	56	37.8	687	1	US-08-583- Sequence 33, Applicati	1.08e+02
28	56	37.8	687	1	US-08-164- Sequence 31, Applicati	1.08e+02
29	56	37.8	687	1	US-08-164- Sequence 31, Applicati	1.08e+02
30	56	37.8	687	1	US-08-583- Sequence 31, Applicati	1.08e+02
31	56	37.8	688	1	US-08-164- Sequence 72, Applicati	1.08e+02
32	56	37.8	688	1	US-08-164- Sequence 70, Applicati	1.08e+02
33	56	37.8	688	1	US-08-583- Sequence 72, Applicati	1.08e+02
34	56	37.8	688	1	US-08-583- Sequence 72, Applicati	1.08e+02
35	56	37.8	722	3	PCT-US95-1 Sequence 2, Applicatio	1.08e+02
36	56	37.8	1651	2	US-08-447- Sequence 2, Applicatio	1.08e+02
37	55	37.2	132	1	US-08-118- Sequence 6, Applicatio	1.33e+02
38	55	37.2	205	2	US-08-684- Sequence 6, Applicatio	1.33e+02
39	55	37.2	301	2	US-08-355- Sequence 2, Applicatio	1.33e+02
40	55	37.2	301	3	PCT-US95-1 Sequence 2, Applicatio	1.33e+02
41	54	36.5	47	3	PCT-US95-0 Sequence 6, Applicatio	1.63e+02
42	54	36.5	263	2	US-08-664- Sequence 1, Applicatio	1.63e+02
43	54	36.5	566	1	US-08-726- Sequence 2, Applicatio	1.63e+02
44	54	36.5	695	1	US-08-583- Sequence 8, Applicatio	1.63e+02
45	54	36.5	3898	2	US-09-059- Sequence 2, Applicatio	1.63e+02

# ALIGNMENTS

RESULT 1  
ID US-08-759-581B-4 STANDARD: PRT: 387 AA.  
AC xxxxxx  
XX  
XX  
XX  
DE Sequence 4, Application US/08759581B  
Sequence 4, Application US/08759581B  
Patent No. 5876945  
GENERAL INFORMATION:  
CC APPLICANT: CHISHOLM, DEXTER A.  
CC APPLICANT: DINER, BRUCE A.  
CC APPLICANT: DONALDSON, GAIL K.  
CC APPLICANT: HERSHEY, HOWARD P.  
CC APPLICANT: JORDAN, DOUGLAS B.  
CC APPLICANT: TANG, XIAO-SONG  
CC APPLICANT: TROST, JEFFREY T.  
CC APPLICANT: WANG, SHAOLIE  
CC APPLICANT: WARREN, PATRICK V.  
CC TITLE OF INVENTION: METHODS FOR IDENTIFYINGHERBICIDAL AGENTS THAT INHIBIT D  
CC NUMBER OF SEQUENCES: 29  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: E.I. DU PONT DE NEMOURS AND COMPANY  
CC STREET: 1007 MARKET STREET  
CC CITY: WILMINGTON  
CC STATE: DELAWARE  
CC COUNTRY: U.S.A.  
CC ZIP: 19898  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: DISKETTE, 3.50 INCH  
CC COMPUTER: IBM PC COMPATIBLE  
CC OPERATING SYSTEM: MICROSOFT WINDOWS 3.1  
CC SOFTWARE: MICROSOFT WORD 2.0C  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/759,581B  
CC FILING DATE:  
CC CLASSIFICATION: 435  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: FLOYD, LINDA AXAMETHY  
CC REGISTRATION NUMBER: 33,692  
CC REFERENCE/DOCKET NUMBER: CR-9964  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: 302-892-8112  
CC TELEFAX: 302-773-0164  
CC INFORMATION FOR SEQ ID NO: 4:

```

CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 387 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: unknown
CC TOPOLOGY: unknown
CC MOLECULE TYPE: protein
CC ORIGINAL SOURCE:
CC STRAIN: SCENESMUS D1 PROTEASE PROTEIN
CC SEQUENCE 387 AA; 40578 MW; 751563 CN;

Query Match 44.6%; Score 66; DB 2; Length 387;
Best Local Similarity 36.8%; Pred. No. 1.25e+01;
Matches 7; Conservative 7; Mismatches 5; Indels 0; Gaps 0;

Dd 2 TSEQLLFEAMRAVDRAVY 20
:::|:|||||:|:|:
Oy 1 ANDHLSILEAMSDNDTPYL 19

LT 2
US-08-759-581B-22 STANDARD; PRT; 464 AA.
xx xxxxxx
xx
xx
xx
xx Sequence 22, Application US/08759581B
DE
Sequence 22, Application US/08759581B
CC Patent No. 5876945
CC GENERAL INFORMATION:
CC APPLICANT: CHISHOLM, DEXTER A.
CC APPLICANT: DINER, BRUCE A.
CC APPLICANT: DONALDSON, GAIL R.
CC APPLICANT: HERSHEY, HOWARD P.
CC APPLICANT: JORDAN, DOUGLAS B.
CC APPLICANT: TANG, XIAO-SONG
CC APPLICANT: TROST, JEFFREY T.
CC APPLICANT: WANG, SHAOJIE
CC APPLICANT: WARREN, PATRICK V.
CC TITLE OF INVENTION: METHODS FOR IDENTIFYINGHERBICIDAL AGENTS THAT INHIBIT D1 F
CC NUMBER OF SEQUENCES: 29
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: E.I. DU PONT DE NEMOURS AND COMPANY
CC STREET: 1007 MARKET STREET
CC CITY: WILMINGTON
CC STATE: DELAWARE
CC COUNTRY: U.S.A.
CC ZIP: 19898
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: DISKETTE, 3.50 INCH
CC COMPUTER: IBM PC COMPATIBLE
CC OPERATING SYSTEM: MICROSOFT WINDOWS 3.1
CC SOFTWARE: MICROSOFT WORD 2.0C
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/759,581B
CC FILING DATE:
CC CLASSIFICATION: 435
CC ATTORNEY/AGENT INFORMATION:
CC NAME: FLOYD, LINDA AXAMETHY
CC REGISTRATION NUMBER: 33,692
CC REFERENCE/DOCKET NUMBER: CR-9964
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 302-892-8112
CC TELEFAX: 302-773-0164
CC INFORMATION FOR SEQ ID NO: 22:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 464 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: unknown
CC TOPOLOGY: unknown
CC MOLECULE TYPE: peptide
CC SEQUENCE 464 AA; 48755 MW; 1062471 CN;

```

```

Query Match          44.6%   Score 66; DB 2; Length 464;
Best Local Similarity 36.8%; Pred. No. 1.25e+01;
Matches              7; Conservative 7; Mismatches 5; Indels 0; Gaps 0;

Db      79 TSEQLFLEAMRAVDRAVY 97
Oy      1 ANDHLSILEAWSMDTPTL 19
        :::|::||| |::|:

RESULT    3
ID US-08-759-581B-9 STANDARD; PRT; 388 AA.
XX AC xxxxxx
XX DT
XX DE
Sequence 9, Application US/08759581B
CC CC Patent No. 5876945
CC CC GENERAL INFORMATION:
CC CC APPLICANT: CHISHOLM, DEXTER A.
CC CC APPLICANT: DINER, BROCE A.
CC CC APPLICANT: DONALDSON, GAIL K.
CC CC APPLICANT: HERSHEY, HOWARD P.
CC CC APPLICANT: JORDAN, DOUGLAS B.
CC CC APPLICANT: TANG, XIAO-SONG
CC CC APPLICANT: TRIST, JEFFREY T.
CC CC APPLICANT: WANG, SHAOJIE
CC CC APPLICANT: WARREN, PATRICK V.
CC CC TITLE OF INVENTION: METHODS FOR IDENTIFYINGHERBICIDAL AGENTS THAT INHIBIT D
CC CC NUMBER OF SEQUENCES: 29
CC CC CORRESPONDENCE ADDRESS:
CC CC ADDRESSEE: E.I. DU PONT DE NEMOURS AND COMPANY
CC CC STREET: 1007 MARKET STREET
CC CC CITY: WILMINGTON
CC CC STATE: DELAWARE
CC CC COUNTRY: U.S.A.
CC CC ZIP: 19898
CC CC COMPUTER READABLE FORM:
CC CC MEDIUM TYPE: DISKETTE, 3.50 INCH
CC CC COMPUTER: IBM PC COMPATIBLE
CC CC OPERATING SYSTEM: MICROSOFT WINDOWS 3.1
CC CC SOFTWARE: MICROSOFT WORD 2.0C
CC CC CURRENT APPLICATION DATA:
CC CC APPLICATION NUMBER: US/08/759,581B
CC CC FILING DATE:
CC CC CLASSIFICATION: 435
CC CC ATTORNEY/AGENT INFORMATION:
CC CC NAME: FLOYD, LINDA AXAMEETHY
CC CC REGISTRATION NUMBER: 33,692
CC CC REFERENCE/DOCKET NUMBER: CR-9964
CC CC TELECOMMUNICATION INFORMATION:
CC CC TELEPHONE: 302-773-0164
CC CC TELEFAX: 302-892-8112
CC CC INFORMATION FOR SEQ ID NO: 9:
CC CC SEQUENCE CHARACTERISTICS:
CC CC LENGTH: 388 amino acids
CC CC TYPE: amino acid
CC CC STRANDEDNESS: unknown
CC CC TOPOLOGY: unknown
CC CC MOLECULE TYPE: protein
CC CC ORIGINAL SOURCE:
CC CC STRAIN: WHEAT DI PROTEASE PROTEIN
CC CC SEQUENCE 388 AA; 41958 MW; 750252 CN;

Query Match          42.6%   Score 63; DB 2; Length 388;
Best Local Similarity 38.9%; Pred. No. 2.41e+01;
Matches              7; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

Db      2 TEENLFLEAMRAVDRAVY 19
        :::|::||| |::|:

```



OY 1 ANDHLSLEAWSNDNTPYL 18

RESULT 4  
ID US-07-857-224B-42 STANDARD: PRT; 267 AA.  
XX xxxxxx

Sequence 42, Application US/07857224B  
Patent No. 5958784  
GENERAL INFORMATION:  
APPLICANT: Benner, Steven A.  
TITLE OF INVENTION: Predicting Folded Structures of Proteins  
NUMBER OF SEQUENCES: 114  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Steven A. Benner  
STREET: Hadlaubstrasse 151  
CITY: Zurich  
STATE: none  
COUNTRY: Switzerland  
ZIP: (note: this is an international post code) CH-8092

COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5 inch diskette, 1.4 MB storage  
COMPUTER: Apple Macintosh  
OPERATING SYSTEM: Macintosh 7.0  
SOFTWARE: Microsoft Word  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/857,224B  
FILING DATE: 03/25/92  
CLASSIFICATION: 436  
PRIOR APPLICATION DATA: none  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (international) 41 1 632 2830  
TELEFAX: (international) 41 1 262 2437  
TELEX: none  
INFORMATION FOR SEQ ID NO: 42:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 267  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
DESCRIPTION: protein  
ORIGINAL SOURCE:  
ORGANISM: Saccharomyces cerevisiae  
FEATURE: Protein kinase; Table 8 Column 48  
PUBLICATION INFORMATION:  
AUTHORS:  
AUTHORS: Hanks, S. K.  
AUTHORS: Quinn, A. M.  
AUTHORS: Hunter, T.  
JOURNAL: Science  
VOLUME: 241  
PAGES: 42-52  
DATE: 1988  
SEQUENCE 267 AA; 30518 MW; 382200 CN;

Query Match 41.9%; Score 62; DB 2; Length 267;  
Best Local Similarity 30.0%; Pred. No. 3.00e+01;  
Matches 6; Conservative 7; Mismatches 7; Indels 0; Gaps 0;

DB 68 TKDNLAIYTWCEGSSLYKH 87  
OY 1 ANDHLSLEAWSNDNTPYL 20

RESULT 5  
ID US-08-909-984A-12 STANDARD: PRT; 315 AA.  
XX xxxxxx

AC xxxxxx  
XX  
DT  
DE  
XX  
CC Sequence 12, Application US/08909984A  
CC Patent No. 5747275  
CC GENERAL INFORMATION:  
CC APPLICANT: Rubin, Gerry M.  
CC APPLICANT: Therrien, Marc  
CC APPLICANT: Chang, Henry C.  
CC APPLICANT: Karim, Felix D.  
CC APPLICANT: Wasserman, David A.  
CC TITLE OF INVENTION: A No. 5747275e1 Protein Kinase Required for Ras  
CC NUMBER OF SEQUENCES: 12  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP  
CC STREET: 268 BUSH STREET, SUITE 3200  
CC CITY: SAN FRANCISCO  
CC STATE: CALIFORNIA  
CC COUNTRY: USA  
CC ZIP: 94104

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/909,984A  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: OSMAN, RICHARD A.  
REGISTRATION NUMBER: 36,627  
REFERENCE/DOCKET NUMBER: B96-010  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 343-4341  
TELEFAX: (415) 343-4342  
INFORMATION FOR SEQ ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 315 amino acids  
TYPE: amino acid  
TOPOLOGY: not relevant  
STRANDEDNESS: not relevant  
MOLECULE TYPE: peptide  
SEQUENCE 315 AA; 35949 MW; 526392 CN;

Query Match 41.9%; Score 62; DB 1; Length 315;  
Best Local Similarity 30.0%; Pred. No. 3.00e+01;  
Matches 6; Conservative 7; Mismatches 7; Indels 0; Gaps 0;

DB 80 TKDNLAIYTWCEGSSLYKH 99  
OY 1 ANDHLSLEAWSNDNTPYL 20

RESULT 6  
ID US-08-909-983-12 STANDARD: PRT; 315 AA.  
XX xxxxxx

Sequence 12, Application US/08909983  
Patent No. 5747288  
GENERAL INFORMATION:  
APPLICANT: Rubin, Gerry M.  
APPLICANT: Therrien, Marc  
APPLICANT: Chang, Henry C.



```
CC TELCOMMUNICATION INFORMATION:  
CC TELEPHONE: (415) 843-5481  
CC TELEFAX: (415) 857-0663  
CC INFORMATION FOR SEQ ID NO: 5:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 346 amino acids  
CC TYPE: amino acid  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: peptide  
CC HYPOTHEICAL: NO  
CC FRAGMENT TYPE: C-terminal  
CC ORIGINAL SOURCE:  
CC ORGANISM: Homo sapiens  
SQ SEQUENCE 346 AA; 39386 MW; 635611 CN;  
  
Query Match 41.9%; Score 62; DB 1; Length 346;  
Best Local Similarity 30.0%; Pred.No. 3.00e+01;  
Matches 6; Conservative 7; Mismatches 7; Indels 0; Gaps 0;  
  
Db 111 TKDNLAIYTCWCEGSSLYKH 130  
:::|:|: : : | |  
Q 1 ANDHLSTLEAWSMDNDTPYLH 20  
  
RESULT 9  
ID US-08-185-282-12 STANDARD: PRT: 648 AA.  
XX xxxxxx  
DT  
DE Sequence 12, Application US/08185282  
XX  
CC Sequence 12, Application US/08185282  
CC Patent No. 5618670  
CC GENERAL INFORMATION:  
CC APPLICANT: Rapp, Ulf R.  
CC APPLICANT: Storm, Stephen M.  
CC TITLE OF INVENTION: DETECTION METHOD FOR C-RAF-1 GENES  
CC NUMBER OF SEQUENCES: 12  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: CUSHMAN,DARBY & CUSHMAN  
CC STREET: 1615 L Street, N.W.  
CC CITY: Washington  
CC STATE: D.C.  
CC COUNTRY: USA  
CC ZIP: 20036-5601  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: PatentIn Release #1.0, Version #1.25  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/185,282  
CC FILING DATE:  
CC CLASSIFICATION: 435  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US/07/759,738  
CC FILING DATE:  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Scott, Watson T.  
CC REGISTRATION NUMBER: 26,581  
CC REFERENCE/DOCKET NUMBER: WTS/5663/82732  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: (202) 861-3000  
CC TELEFAX: (202) 822-0944  
CC TELEX: 67146Z7 CUSH  
CC INFORMATION FOR SEQ ID NO: 12:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 648 amino acids  
CC TYPE: amino acid  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: protein  
CC
```

SO	SEQUENCE	648 AA; 73051 MW; 2254837 CN;	
	Query Match	41.9%; Score 62; DB 1; Length 648;	
	Best Local Similarity	30.0%; Pred. No. 3.00e+01;	
	Matches	6; Conservative	7; Mismatches 7; Indels 0; Gaps 0;
Db	413 TKDNIAIVTQWCGSSSLYKH 432		
	1 ANDHLSILEAWSMDNDTPYLH 20		
QY	1 ANDHLSILEAWSMDNDTPYLH 20		
RESULT	10		
ID	US-08-886-751A-6	STANDARD;	PRT; 648 AA.
XX	xxxxxx		
DE	Sequence 6, Application US/08886751A		
XX	Sequence 6, Application US/08886751A		
CC	Patent No. 5885783		
CC	GENERAL INFORMATION:		
CC	APPLICANT: YOO, Tai-June		
CC	APPLICANT: Cheng, Kuang-Chuan		
CC	TITLE OF INVENTION: Autoimmune Inner Ear Disease Antigen and		
CC	TITLE OF INVENTION: Diagnostic Assay		
CC	NUMBER OF SEQUENCES: 6		
CC	CORRESPONDENCE ADDRESS:		
CC	ADDRESSEE: Jones & Askew		
CC	STREET: 191 Peachtree Street, 37th Floor		
CC	CITY: Atlanta		
CC	STATE: Georgia		
CC	COUNTRY: U.S.A.		
CC	ZIP: 30303		
CC	COMPUTER READABLE FORM:		
CC	MEDIUM TYPE: Floppy disk		
CC	COMPUTER: IBM PC compatible		
CC	OPERATING SYSTEM: PC-DOS/MS-DOS		
CC	SOFTWARE: Patentin Release #1.0, Version #1.30		
CC	CURRENT APPLICATION DATA:		
CC	APPLICATION NUMBER: US/08/886,751A		
CC	FILING DATE:		
CC	CLASSIFICATION: 424		
CC	ATTORNEY/AGENT INFORMATION:		
CC	NAME: Young, Leona G.		
CC	REGISTRATION NUMBER: 37,266		
CC	REFERENCE/DOCKET NUMBER: 25490-0100		
CC	TELECOMMUNICATION INFORMATION:		
CC	TELEPHONE: (404) 818-3700		
CC	TELEFAX: (404) 818-3799		
CC	INFORMATION FOR SEQ ID NO: 6:		
CC	SEQUENCE CHARACTERISTICS:		
CC	LENGTH: 648 amino acids		
CC	TYPE: amino acid		
CC	STRANDEDNESS: single		
CC	TOPOLOGY: linear		
CC	MOLECULE TYPE: protein		
CC	HYPOTHETICAL: NO		
CC	ANTI-SENSE: NO		
CC	FRAGMENT TYPE: N-terminal		
CC	ORIGINAL SOURCE:		
CC	ORGANISM: Homo sapiens		
CC	SEQUENCE	648 AA; 73051 MW; 2254837 CN;	
Db	Query Match	41.9%; Score 62; DB 2; Length 648;	
	Best Local Similarity	30.0%; Pred. No. 3.00e+01;	
	Matches	6; Conservative	7; Mismatches 7; Indels 0; Gaps 0;
QY	1 ANDHLSILEAWSMDNDTPYLH 20		

RESULT 11  
ID US-08-276-151-2 STANDARD: PRT: 648 AA.  
XX  
AC xxxxxx  
XX  
DE Sequence 2, Application US/08276151  
XX  
CC Sequence 2, Application US/08276151  
CC Patent No. 5597719  
CC GENERAL INFORMATION:  
CC APPLICANT: Freed, Ellen  
CC APPLICANT: Ruggieri, Rosamaria  
CC TITLE OF INVENTION: Interaction of raf-1 and 14-3-3 Proteins  
CC NUMBER OF SEQUENCES: 9  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: Cooley Godward et al.  
CC STREET: Five Palo Alto Square  
CC CITY: Palo Alto  
CC STATE: CA  
CC COUNTRY: USA  
CC ZIP: 94036  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: Patentin Release #1.0, Version #1.25  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/276,151  
CC FILING DATE: 14-JUL-1994  
CC CLASSIFICATION: 530  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Torchia, Ph.D., Timothy E  
CC REGISTRATION NUMBER: 36,700  
CC REFERENCE/DOCKET NUMBER: ONIX-005/0005  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: (415) 843-5481  
CC TELEFAX: (415) 857-0663  
CC INFORMATION FOR SEQ ID NO: 2:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 648 amino acids  
CC TYPE: amino acid  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: protein  
CC SEQUENCE 648 AA; 73051 MW; 2254837 CN;  
Query Match 41.9%; Score 62; DB 1; Length 648;  
Best Local Similarity 30.0%; Pred. No. 3.00e+01;  
Matches 6; Conservative 7; Mismatches 7; Indels 0; Gaps 0;

Db 413 TKDNLAIWTMCSSSLYKH 432  
QY 1 ANDHLSTLEAMSDNDPYLH 20

RESULT 12  
ID US-08-185-282-3 STANDARD: PRT: 648 AA.  
XX  
AC xxxxxx  
XX  
DE Sequence 3, Application US/08185282  
XX  
CC Sequence 3, Application US/08185282  
CC Patent No. 5618670  
CC GENERAL INFORMATION:  
CC APPLICANT: Rapp, Ulf R.  
CC APPLICANT: Storm, Stephen M.  
CC TITLE OF INVENTION: DETECTION METHOD FOR C-RAF-1 GENES  
CC NUMBER OF SEQUENCES: 12

CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: CUSHMAN, DARBY & CUSHMAN  
CC STREET: 1615 L Street, N.W.  
CC CITY: Washington  
CC STATE: D.C.  
CC COUNTRY: USA  
CC ZIP: 20036-5601  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: Patentin Release #1.0, Version #1.25  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/185,282  
CC FILING DATE:  
CC CLASSIFICATION: 435  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US/07/759,738  
CC FILING DATE:  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Scott, Watson T.  
CC REGISTRATION NUMBER: 26,581  
CC REFERENCE/DOCKET NUMBER: KTS/5683/82732  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: (202) 861-3000  
CC TELEFAX: (202) 822-0944  
CC TELEEX: 6714627 CUSH  
CC INFORMATION FOR SEQ ID NO: 3:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 648 amino acids  
CC TYPE: amino acid  
CC TOPOLOGY: linear  
CC SEQUENCE 648 AA; 72915 MW; 2223546 CN;  
SQ

Query Match 41.9%; Score 62; DB 1; Length 648;  
Best Local Similarity 30.0%; Pred. No. 3.00e+01;  
Matches 6; Conservative 7; Mismatches 7; Indels 0; Gaps 0;

Db 413 TKDNLAIWTMCSSSLYKH 432  
QY 1 ANDHLSTLEAMSDNDPYLH 20

RESULT 13  
ID US-08-185-282-1 STANDARD: PRT: 648 AA.  
XX  
AC xxxxxx  
XX  
DE Sequence 1, Application US/08185282  
XX  
CC Sequence 1, Application US/08185282  
CC Patent No. 5618670  
CC GENERAL INFORMATION:  
CC APPLICANT: Rapp, Ulf R.  
CC APPLICANT: Storm, Stephen M.  
CC TITLE OF INVENTION: DETECTION METHOD FOR C-RAF-1 GENES  
CC NUMBER OF SEQUENCES: 12  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: CUSHMAN, DARBY & CUSHMAN  
CC STREET: 1615 L Street, N.W.  
CC CITY: Washington  
CC STATE: D.C.  
CC COUNTRY: USA  
CC ZIP: 20036-5601  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: Patentin Release #1.0, Version #1.25  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/185,282

CC 1 ELEV: 0/14027 CUSH  
CC INFORMATION FOR SEQ ID NO: 4:

QY 1 ANDHLSILEAWSNDNTPYLH 20

Thu Jan 13 09:13:31 2000

US-09-290-049-1.rai

Search completed: Tue Jan 11 15:39:05 2000  
Job time : 8 secs.